A Publication of the American Psychological Association

The Journal of Abnormal and Social Psychology

Founded by Morton Prince, 1906

CONTENTS

ARTICLES

Ideology, Personality, and Institutional Policy in the Mental Hospital 263
Doris C. Gilbert and Daniel J. Levinson

Personality Structure and Group Structure: An Interpretative Study of Their Relationship Through an Event-Structure Hypothesis 272
Arnold S. Tannenbaum and Floyd H. Allport

Stimulus Generalization as a Function of Clinical Anxiety 281
Gerald Rosenbaum

Distractibility in the Conceptual Performance of Schizophrenics 286
Loren J. Chapman

Scaling Norm Conflicts in the Area of Prejudice and Discrimination 292
Nathan Kogan and James F. Downey

Development and Evaluation of a Scale for Measuring Social Acquiescence ... 296
Bernard M. Bass

Pride in Group Performance and Group-Task Motivation 300
Leonard Berkowitz and Bernard I. Levy

Continued on inside front cover

This is the Last Issue of Volume 53.
Title Page and Contents for the Volume
Appear Herein.

Volume 53 NOVEMBER 1956 Number 3

CONTENTS – *Continued*

Effects of Work-Group Structure and Certain Task Variables on Group Performance . . . 307
John T. Lanzetta and Thornton B. Roby
Effects of Glutamic Acid and Social Stimulation in Mental Deficiency . . . 315
Guinevere S. Chambers and Ralph N. Zabarenko
Retention of Affectively Toned Verbal Material by Normals and Neurotics . . . 321
Samuel F. Klugman
Reminiscence, Drive, and Personality Theory . . . 328
H. J. Eysenck
Perceptual Organization and Intelligence: A Further Study . . . 334
Clinton DeSoto and H. Leibowitz
Impressions of Personality, Authoritarianism, and the *Fait-Accompli* Effect . . . 338
I. H. Paul
Relations of Temperament to the Choice of Values . . . 345
Lyle V. Jones and Charles Morris
Enhancement of Contrast as Composite Habit . . . 350
Donald T. Campbell
A Factorial Study of Dogmatism and Related Concepts . . . 356
Milton Rokeach and Benjamin Fruchter
Performance as a Function of Motive Strength and Expectancy of Goal-Attainment . . . 361
John W. Atkinson and Walter R. Reitman
Cognitive Structure and Attitudinal Affect . . . 367
Milton J. Rosenberg
The Effects of Emotional Arousal on the Retention of Film Content: A Failure to Replicate . . . 373
Eleanor E. Maccoby, Harry Levin, and Bruce M. Selya

CASE REPORT

The Response of Schizophrenic Patients to a Televised World Series Game: A Study in Social Isolation . . . 375
Jerome L. Singer and Herbert E. Spohn

CRITIQUE AND NOTES

Authoritarianism and Interpersonal Perception . . . 378
W. H. Crockett and Thomas Meidinger
Ethnic Cleavage Among Young Children . . . 380
Wallace E. Lambert and Yoshi Taguchi
Differences in the Perception of an Authority Figure and a Nonauthority Figure by Navy Recruits . . . 383
Warren J. Wittreich and Keith B. Radcliffe, Jr.
A Note on the Social Perceptions of Authoritarians and Nonauthoritarians . . . 384
William Rabinowitz
A Comparison of the DRQ With Ratings of Emotion . . . 386
Frank Auld, Jr., and George F. Mahl
A Simple Device for Obtaining Certain Verbal Activity Measures During Interviews . . . 388
Stanislav V. Kasl and George F. Mahl

Volume 53 NOVEMBER 1956 Number 3

The Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, New York University

Associate Editor: E. J. SHOBEN, JR., Teachers College, Columbia University

Managing Editor: LORRAINE BOUTHILET

CONSULTING EDITORS

GORDON W. ALLPORT
Harvard University

URIE BRONFENBRENNER
Cornell University

DONALD T. CAMPBELL
Northwestern University

LAUNOR F. CARTER
The RAND Corporation

RICHARD CHRISTIE
Columbia University

MORTON DEUTSCH
Bell Telephone Laboratories

I. E. FARBER
State University of Iowa

LEON FESTINGER
Stanford University

DONALD W. FISKE
University of Chicago

ROBERT R. HOLT
New York University

HERBERT H. HYMAN
Columbia University

IRVING L. JANIS
Yale University

HAROLD E. JONES
University of California

HAROLD H. KELLEY
Center for Advanced Study in the Behavioral Sciences

GARDNER LINDZEY
Syracuse University

DAVID C. MCCLELLAND
Harvard University

QUINN MCNEMAR
Stanford University

LEO POSTMAN
University of California

NEVITT SANFORD
Vassar College

SEYMOUR SARASON
Yale University

RICHARD L. SOLOMON
Harvard University

RENATO TAGIURI
Harvard University

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

MT. ROYAL & GUILFORD AVES., BALTIMORE 2, MD.
AND 1333 SIXTEENTH STREET N.W., WASHINGTON 6, D. C.

Entered as second-class matter February 15, 1954 at the Post Office at Baltimore, Maryland under the Act of February 28, 1925. Acceptance for mailing at special rate of postage provided for in Section 34.40 (e), P. L. and R. Additional entry at Washington, D.C.

Copyright 1956 by the American Psychological Association, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis; those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area, this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued bimonthly in January, March, May, July, September, and November, two volumes per year. The two volumes average 784 pages. The subcription price per year is $16.00; foreign, $16.50. Single issues are $3.00.

Original manuscripts should be sent to the Editor:

DR. M. BREWSTER SMITH
Graduate Department of Psychology
New York University
Washington Square
New York 3, N. Y.

Fifty free offprints are supplied to contributors. Authors of prior publication articles receive no gratis offprints.

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Printed and mailed by:

WAVERLY PRESS, INC.
Baltimore 2, Maryland

IDEOLOGY, PERSONALITY, AND INSTITUTIONAL POLICY IN THE MENTAL HOSPITAL[1]

DORIS C. GILBERT AND DANIEL J. LEVINSON

Harvard University

THIS inquiry concerns the ideologies of mental hospital members regarding the nature and causes of mental illness and regarding hospital aims and policies in treating the mentally ill. We are interested in the nature of these ideologies as well as their institutional and intrapersonal roots. Mental hospitals are going through a period of ideological ferment and organizational change. The newly emerging viewpoints involve much more than the application of new treatment techniques. What gives them their fundamental quality is their conception of the hospital as a community of citizens rather than a rigidly codified institutional mold, and their conception of the hospital members as persons rather than as mere objects and agents of treatment (2, 8, 9, 14). In this respect as in many others, the ongoing developments in the mental hospital parallel those in other "correctional" institutions such as schools and prisons, and in larger bureaucratic structures such as industry and government (10, 12, 15). Because of their wider relevance, research in the mental hospital can make use of, and contribute materially to, the main body of sociopsychological theory and knowledge.

The primary aims of this study are the following:

1. To formulate the main characteristics of the old and the newly emerging viewpoints regarding mental illness, and to construct an ideology scale that will crudely measure the degree of an individual's preference for one or the other viewpoint.

2. To investigate the personality contexts within which these orientations most readily develop.

3. To investigate the relationships of individual ideology and personality to membership in particular types of hospital systems and occupational statuses.

4. To investigate the ways in which the hospital's over-all policy is related to the modal (most common) ideology and the modal personality of its members.

[1] This investigation was supported in part by research grants M-687 and M-1000 from the National Institute of Mental Health, of the National Institutes of Health, Public Health Service.

CUSTODIALISM AND HUMANISM AS IDEOLOGICAL ORIENTATIONS IN THE MENTAL HOSPITAL

We have found it most fruitful, in our attempts at ideological analysis, to begin by asking: What are the major "problems" or issues of institutional life for which some kind of adaptive rationale is needed? The mental hospital presents at least the following issues with which every individual member must deal. The *patient*: what is he like, how did he get that way, how much can he be helped and in what ways? *Patient-staff relations:* what should be the role of hospital personnel vis-à-vis the patient; how much interaction should there be, with what emotional qualities and therapeutic aims? *Staff-staff relations:* how should specific functions and responsibilities be distributed; how much communication and status distinction should there be? *General hospital practices:* what formal treatment methods are best; how should ward life be organized; in what ways should patients be encouraged, left alone, controlled, punished?

The traditional and the newly developing viewpoints offer two contrasting sets of answers to the above questions. Each is a mode of thought underlying institutional policy and individual adaptation. A brief, schematic formulation of each ideology is presented below. It is to be emphasized that these are prototypes, probably represented in pure form by few individuals or hospitals. The two types may be thought of as polar extremes of a continuum containing various intermediate positions. We assume that this continuum is *realistic*, in the sense that viewpoints and policies approximating the prototypes will commonly be found, and that the continuum is *significant*, in the sense that adherence to one rather than the other viewpoint will seriously affect the individual's or the hospital's mode of functioning.

We shall use the term "custodialism" to

designate the traditional patterns. The model of the custodial orientations is the traditional prison and the "chronic" mental hospital which provide a highly controlled setting concerned mainly with the detention and safekeeping of its inmates. Patients are conceived of in stereotyped terms as categorically different from "normal" people, as totally irrational, insensitive to others, unpredictable, and dangerous. Mental illness is attributed primarily to poor heredity, organic lesion, and the like. In consequence, the staff cannot expect to understand the patients, to engage in meaningful relationships with them, nor in most cases to do them much good. Custodialism is saturated with pessimism, impersonalness, and watchful mistrust. The custodial conception of the hospital is autocratic, involving as it does a rigid status hierarchy, a unilateral downward flow of power, and minimal communication within and across status lines.

The newer orientations will be termed "humanistic" (after Fromm (6)) in view of their concern with the individuality and the human needs of both patients and personnel. These orientations conceive of the hospital as a therapeutic community rather than a custodial institution. They emphasize interpersonal and intrapsychic sources of mental illness, often to the neglect of possible hereditary and somatic sources. They view patients in more psychological and less moralistic terms. They are optimistic, sometimes to an unrealistic degree, about the possibilities of patient recovery in a maximally therapeutic environment. They attempt in varying degrees to democratize the hospital, to maximize the therapeutic functions of nonmedical personnel, to increase patient self-determination individually and collectively, and to open up communication wherever possible. While the humanistic orientations have the above characteristics in common, and even more an opposition to custodialism, they still differ among themselves in important respects. For example, the concrete manifestations of humanism will differ, although the guiding spirit may be the same, in a large, architecturally horrendous, financially limited state hospital, as contrasted with a small, well-subsidized, private hospital that accepts only patients regarded as good therapeutic risks.

This inquiry makes no assumptions about the actual therapeutic effectiveness of the various approaches. Our primary concern here is with the nature and the determinants of custodialism and humanism.

THE CUSTODIAL MENTAL ILLNESS IDEOLOGY SCALE (CMI)

The initial, field exploration led to the formulation of the "custodialism-humanism" continuum, the polar extremes of which have been described above. The next step was to construct the Custodial Mental Illness Ideology Scale (CMI), an admittedly crude instrument that had two chief functions in the research: (*a*) To test the hypothesis that a set of seemingly disparate ideas do in fact "go together" to form a relatively coherent orientation in the individual. A derivative function is to determine whether viewpoints approximating our posited prototypes exist with some frequency within various hospital settings. (*b*) To provide a quantitative and at the same time meaningful measure that facilitates additional study of the nature of these ideas and their relation to other aspects of the individual and his milieu.

The CMI scale consists of 20 statements, broadly diversified to cover numerous facets of the ideological domain: the nature and causes of mental illness, conditions in the hospital, patient-staff relations, and the like. The items were derived from interviews, conversations, and observations of conferences and everyday hospital life. The scale is presented in Table 1. A more extensive description of the field work and derivation of the CMI scale can be found in Gilbert (7).

Scoring procedure. The subjects were instructed to indicate the degree of their agreement or disagreement with each item on a scale ranging from +3 (strong agreement) to −3 (strong disagreement). The responses were converted into scores by means of an a priori, 7-point scoring scheme. It was intended that a high score represent strong adherence to "custodial" ideology as here conceived, and that a low score represent opposition to this viewpoint. Of the 20 scale items, 17 were regarded as custodial, 3 as humanistic. For the "custodial" items, seven points were given for the +3 response, one point for −3. For the "humanistic" items the

TABLE 1

THE CUSTODIAL MENTAL ILLNESS IDEOLOGY (CMI) SCALE

Item	Mean	*DP*
1. Only persons with considerable psychiatric training should be allowed to form close relationships with patients.	3.5	2.4
3. It is best to prevent the more disturbed patients from mixing with those who are less sick.	5.0	1.8
5. As soon as a person shows signs of mental disturbance he should be hospitalized.	3.3	4.2
*7. Mental illness is an illness like any other.	2.7	2.7
9. Close association with mentally ill people is liable to make even a normal person break down.	2.0	1.4
11. We can make some improvements, but by and large the conditions of mental hospital wards are about as good as they can be considering the type of disturbed patient living there.	3.0	3.6
15. We should be sympathetic with mental patients, but we cannot expect to understand their odd behavior.	3.2	3.8
17. One of the main causes in mental illness is lack of moral strength.	2.8	3.2
*18. When a patient is discharged from a hospital, he can be expected to carry out his responsibilities as a citizen.	3.0	.5
19. Abnormal people are ruled by their emotions; normal people by their reason.	3.8	4.4
21. A mental patient is in no position to make decisions about even everyday living problems.	3.0	3.1
*23. Patients are often kept in the hospital long after they are well enough to get along in the community.	4.2	−.2
25. There is something about mentally ill people that makes it easy to tell them from normal people.	3.0	2.9
27. Few, if any, patients are capable of real friendliness.	2.2	1.7
31. There is hardly a mental patient who isn't liable to attack you unless you take extreme precautions.	2.5	3.0
33. Patients who fail to recover have only themselves to blame; in most cases they have just not tried hard enough.	1.8	1.5
37. "Once a schizophrenic, always a schizophrenic."	2.3	1.3
38. Patients need the same kind of control and discipline as an untrained child.	3.3	2.4
39. With few exceptions most patients haven't the ability to tell right from wrong.	2.4	2.3
40. In experimenting with new methods of ward treatment, hospitals must consider, first and foremost, the safety of patients and personnel.	5.3	2.3

* Items expressing a "humanistic" position; all others are "custodial."

Note.—The item means and *DP*s are those obtained by a sample of 196 mental hospital personnel in Hospitals C, T, and H. Similar *DP*s have been obtained in other samples of personnel, patients, and visitors. Items are numbered as they appear in the questionnaire, which contained other scales and questions.

scoring was reversed. For convenience in comparing scores from scales differing in length, we shall use the mean per item, multiplied by 10. The possible range is thus 10–70 points.

The CMI scale was initially developed on a sample of 335 staff members (aides, student nurses, nurses, and psychiatrists) in three Massachusetts mental hospitals: Hospital C, a large (1,800 bed) institution dealing largely with "chronic" patients; Hospital T, a Veterans Administration hospital of about the same size; and Hospital H, a small (120 bed) state institution providing short-term active treatment. The range for this sample was 15–52, the mean being 31.3 and the *SD*, 9.5. Comparative data for various subgroupings are presented below (Table 2). The reliability (split-half correlation, corrected by Spearman-Brown formula) was .85, and test-retest correlations on several small groups were of similar magnitude. Table 1 presents the means and discriminatory powers (*DP*) of the individual items.[2] The *DP*s of all items except numbers 18 and 23 reach the .05 level of statistical significance, and most of them are beyond the .01 level.

The above data indicate that the initial form of the CMI scale has adequate reliability and internal consistency, and they provide a

[2] *DP* of an item reflects its ability to differentiate between extremely high scorers and extreme low scorers (the upper and lower 25 per cent) on the total scale. It is computed as the difference between the means of the high-scoring and the low-scoring groups.

TABLE 2
CMI MEAN, F MEAN, AND INDEX OF STATUS-CUSTODIALISM

Hospital-Status Unit	*N*	Index of Status-Custodialism	CMI Scale			F Scale			$r_{CMI\,F}$
			Mean	Rank	*SD*	Mean	Rank	*SD*	
Attendants at Hospital:									
C	29	12	38.0	12	10.2	46.1	12	13.0	.91
T	51	11	37.3	11	10.3	41.4	11	11.7	.82
H	48	10	32.1	8	9.7	32.5	9	13.0	.77
Student Nurses at Hospital:									
C	66	9	33.4	10	7.7	29.0	8	10.9	.44
T	38	8	33.3	9	6.8	28.9	7	8.9	.25
H	16	7	31.3	6	5.7	28.8	6	9.2	.59
Nurses at Hospital:									
C	14	6	31.3	7	9.9	37.3	10	14.5	.90
T	18	5	22.4	2	5.2	26.7	5	8.7	.53
H	21	4	26.9	5	7.7	25.8	3	14.5	.73
Doctors at Hospital:									
C	6	3	25.8	4	7.0	18.1	1	6.2	.75
T	4	2	21.6	1	4.8	26.6	4	12.8	.80
H	24	1	22.7	2	4.5	19.1	2	8.5	.50
Total Status:									
Attendants	128	4	35.3	4	10.4	39.1	4	13.7	.82
Student Nurses	120	3	33.1	3	7.3	28.9	3	10.2	.41
Nurses	53	2	26.5	2	8.3	27.9	2	12.8	.76
Doctors	34	1	23.1	1	5.3	19.7	1	9.1	.46
Total Hospital:									
C	115	3	33.7	3	8.9	33.8	2	14.5	.67
T	111	2	32.9	2	10.2	34.2	3	12.4	.69
H	109	1	29.0	1	8.7	27.7	1	13.1	.76
Total Sample	335		31.3		9.5	31.9		13.7	.71

basis for further improvement. They suggest, moreover, that a person's stand on any single issue represented in the scale is part of a broader, fairly coherent (though seldom fully integrated) ideology that embraces numerous issues of hospital life.

Data were obtained on two "validation groups" to determine whether the CMI score adequately gauges an enduring ideological conviction. One group, containing 10 administrators at Hospital H who are known for their advocacy of humanistic policies, earned a CMI mean of 18.8, with an *SD* of 6.1. The second group comprised the professional staff at the Social Rehabilitation Unit, Belmont Hospital, England, (9) and would also be expected to have a low CMI mean. The obtained mean was 22.9, the *SD*, 6.7. These findings offer additional evidence of scale validity.[3]

[3] In a study to be published shortly, we have obtained correlations of .5 and .8 between the CMI scale and a measure of custodialism in role performance, in two samples of hospital aides. Since at least a moderate relation between ideology and action would be expected on theoretical grounds, evidence of such a relation has indirect validational relevance for the CMI scale (4).

PSYCHOLOGICAL BASES OF CUSTODIALISM AND HUMANISM

With the CMI scale developed, it was possible to test the hypothesis that *the custodial orientation is one facet of an authoritarian personality, the humanistic orientation a facet of an equalitarian personality.* Several lines of theory and observation led to this expectation. Custodialism is strongly autocratic in its conception of the hospital and ethnocentric in its conception of patients as an inferior and threatening outgroup entitled to few if any of the rights of "normal" people. Humanism, on the other hand, seeks a more democratic hospital structure and regards patients as individuals to be understood and treated rather than as an outgroup to be pitied or condemned. There is considerable evidence that autocratic viewpoints tend to exist within authoritarian personality structures (1, 5, 6).

We accordingly predicted that the CMI scale would correlate significantly with the F scale (1), a relatively nonideological measure of authoritarianism, and with the Traditional Family Ideology (TFI) Scale (11), a measure of autocratic ideology regarding issues such as husband-wife and parent-child relations.[4]

The obtained interscale correlations are as follows: CMI and F correlate .67, .69, and .76 in Hospitals C, T, and H, respectively. The comparable correlations between CMI and TFI are .50, .56, and .77. The respective *N*s in the three hospitals were 115, 111, and 109. The scale means and *SD*s are presented in Table 2. The findings lend support to the hypothesis that an individual's views regarding mental illness and the hospital are imbedded within a broader ideological and psychodynamic matrix.

It may be argued that the F scale is made of the same stuff as CMI, that it taps relatively superficial ideas or values rather than more central aspects of personality. If this be true, then the foregoing inferences concerning the psychodynamic bases of ideology are unjustified. It is certainly possible that a person may accept many of the ideas represented in the F scale without being an "authoritarian personality." However, we propose on both theoretical and empirical grounds that such persons are rather the exception than the rule. The F items taken as a whole do not comprise an organized body of doctrine. The obtained consistency of response to these items is, we believe, determined for the most part by an enduring pattern of intrapersonal dispositions. Empirical support for this view is given by Adorno, *et al.* (1) and others; for a critical summary, see Christie (3). Significant relationships between CMI scores and nonscale measures of authoritarianism have been obtained by Gilbert (7) and Pine (13). These studies utilized content analysis of interviews, TATs, open-ended questions, and the like in assessing authoritarianism. The F scale would seem to provide a relatively valid though by no means infallible estimate of personal authoritarianism.

[4] An abbreviated F scale of 8 items was used; it contained Items 9, 13, 18, 25, 26, 37, and 42 from Form 45, and Item 32 from Form 78 of the original F scale (1). The TFI measure contained Items 2, 3, 5, 6, 7, 9, 11, 12 of the short form presented by Levinson and Huffman (14, p. 268).

Custodial ideology has important psychic functions for authoritarian hospital members. The idea that patient behavior is simply irrational and not understandable has great value in reducing inner strain and maintaining self-esteem for personnel who have difficulty at the outset in taking an intraceptive, psychological approach. Again, for the person who has a great defensive need to displace and project aggressive wishes concerning authority figures to those who can be regarded as immoral, custodial ideology has special equilibrium-maintaining value through its justification of punitive, suppressive measures.

Humanistic ideology has corresponding functions for its adherents. By supporting a critical attitude toward the established order, it permits many equalitarian individuals to express generalized anti-authority hostilities in an ego-syntonic form. The principle of "self-control through self-understanding," applied in the treatment of patients, often serves to maintain and consolidate the intellectualizing defenses of equalitarian personnel. In our view, then, both custodialism and humanism have important nonrational functions for their proponents.

RELATIONSHIPS AMONG IDEOLOGY, PERSONALITY, AND HOSPITAL POLICY

We have been concerned thus far with ideology as an aspect of personality. We have suggested that the individual's orientation to mental illness is an intrinsic part of his general approach to life problems and is related to deeper-lying personality dynamics. This, however, is only part of the story. Ideology is also an aspect of the social milieu; we must consider both the psychological and the social soil within which ideologies are formed and modified.

Various social factors operate to induce some degree of ideological uniformity among members of a given occupational status, as well as among members of a total hospital system. Many social scientists, including psychologists, argue or implicitly assume that ideological conformity is ordinarily achieved and that some sort of J curve or concentration of viewpoints approximating the institutional requirements will be found among members of a given institution. One serious limitation of this approach, in our view, is its neglect of the

part played by personality. We would expect that the achievement of a policy-congruent modal ideology depends in part on the presence of a corresponding modal personality. Conversely, to the extent that there is variability in ideology-relevant personality characteristics, we would expect ideological variability among members of any system.

Our three domains of inquiry are ideology, personality, and system requirements. Within each domain we have measured individual or system differences along a given continuum: (*a*) The custodial-humanistic continuum of individual ideology, as measured by the CMI scale. (*b*) The authoritarian-equalitarian continuum of personality, as measured by the F scale, (*c*) The third continuum, custodialism-humanism in system requirements, which was assessed as follows. The sample contained 12 subsystems: four occupational statuses (aide, student nurse, nurse, and psychiatrist) in each of three hospitals. Our procedure was to rank the 12 systems in order from relatively most custodial to relatively most humanistic with regard to the demands and pressures each system placed on its members. We ranked the hospitals first, then the statuses, and then combined the two sets into one series of 12 ranks.

The hospitals were ranked in terms of their degree of change away from a predominantly custodial emphasis on protection and bodily care of patients. The large state hospital, C, was assessed as the most custodial in view of its structural emphasis on detention, protection, and custodial care of patients in a highly controlled setting. The pressures it exerted on personnel, and the kinds of experience it offered them, seemed most conducive of a custodial orientation. The large VA hospital, T, was considered intermediate or transitional in that it was in process of fairly rapid change away from custodialism. The third hospital, H, was the most humanistic of the three in its program of ward care, patient government, and general staff-patient relationships.

The four statuses were ranked in degree of custodialism on the basis of educational level and job requirements vis-à-vis the patient. In order from high to low in degree of custodialism, they fall as follows: aide, student nurse, nurse, and psychiatrist.

The three hospitals and the four statuses were then combined into a series of 12 hospital-status units ranked according to degree of pressure toward custodialism. Since occupational status pressures operate over a longer period of time, and more selectively, than do hospital pressures, we made status a primary basis of stratification, and hospital secondary. That is, we assumed that statuses are relatively nonoverlapping in degree of custodialism in their policy requirements and that hospitals make a difference only within a single status grouping. Accordingly the rank 12 was given to the most custodial status in the most custodial hospital, namely, the aide status at Hospital C; this is followed by the aide status at Hospital T, and at H; then come the student nurse statuses at C, T, and H; the nurse statuses at C, T, and H; and lastly the doctor statuses at Hospitals C, T, and H with ranks of 3, 2, and 1 respectively (see Table 2). Ideally a more intensive sociological analysis of the structure and policies of each status in each hospital should be carried out. However, the rankings used here seem adequate for our present purposes.

Having roughly assessed the degree of custodialism in the policy requirements of each hospital-status system, we can now investigate the degree to which these requirements are supported by the ideologies, and are congruent with the personalities, of the system members.

Relations between policy requirements and ideology. What is the relationship between the degree of custodialism in the policy requirements of a hospital-status system and the degree of custodialism in the modal ideology of its members? The relevant data are given in Table 2. We use the CMI mean as a measure of modal ideology, for in the distribution of CMI scores the mean, by and large, corresponds closely to the mode. The obtained rank-order correlation between degree of custodialism in policy requirements (status ranks) and in modal ideology (CMI mean) is .92. There is, in other words, relatively great congruence between policy demands and modal ideology. At the same time, the CMI means of the 12 status units do not correspond fully in absolute degree to the estimated degree of custodialism in their structural pressures. For example, the aide status at Hospital C was ranked most custodial both in policy requirements and in CMI mean; however, in an absolute sense, the policy requirements are

highly custodial whereas the CMI mean is only moderate.

The above findings do not tell us how much ideological variability exists within each system. Data on variability are given in Table 2. It will be noted that the *SD*s of most of the 12 units approximate the *SD* for the total sample. Only in the doctor statuses is there anything approaching uniformity of opinion. Thus although *modal* ideology is fairly closely related to policy requirements, the findings on intrasystem variability suggest that an individual's ideology does not reflect in a simple way the demands of his occupational milieu.

In investigating the relationship between *individual* ideology and policy requirements, we consider system pressures as characteristics of the individual. Every individual in the sample of 335 was assigned an index figure representing the relative degree of custodialism in the policy requirements of his particular hospital-status unit. This index figure is simply the rank of the individual's status within the series of 12. For instance, each doctor at H, the least custodial status, is assigned an index figure of 1, and each aide at C is assigned an index of 12.

The obtained product-moment correlation[5] between CMI score and Index of Status-Custodialism is .47. This finding is evidence of a significant but moderate relationship between individual ideology and system pressures. If system pressures were the most weighty determinants of individual ideology, relative ideological homogeneity within statuses should follow, and thus a high correlation (of the order, .7 to .8) between an individual's CMI score and the degree of policy-custodialism of his work unit. However, the degree of uniformity within any system is not as great as a system-centered mode of thinking would require. An individual's ideology can be predicted with only fair accuracy on the basis of his occupational-hospital membership.

Relations between policy requirements and personality. If the individual's ideological orientation is thought to be simply and directly a result of pressures from his work milieu, relatively independent of his personality, one would not expect the degree of custodialism in system policy to be significantly related to the degree of authoritarianism in modal personality. Rather, the 12 units might be expected to show similar degrees of authoritarianism, as measured by the F-scale means.

In our view, however, some congruence is to be expected between the policy requirements of a system and the modal personality of its members. Such congruence would be facilitated through recruitment, selective turnover, and possible personality changes in the direction of congruence. We are supported in this hypothesis by the finding of congruence (the correlation of .92) between policy requirements and modal ideology in the 12 status units. We would expect a parallel correspondence between policy requirements and modal personality.

For the 12 status units, the obtained rank-order correlation between Index of Status-Custodialism and F mean is .90 (see Table 2). Thus, there is relatively great congruence between policy demands and modal personality. This congruence is as great as that between policy demands and modal ideology.

The obtained correspondence between modal personality (F) and system requirements is accompanied by appreciable variability on the F scale within most of the statuses (Table 2). The size of the variance on F tends to covary with that on CMI ($r = .61$). This leads us to consider the degree to which system membership and personality are related in the individual. We would expect that the correlation found above between status membership and individual CMI score (.47) will hold as well for index of status membership and F score. The findings bear out this prediction. The correlation between the individual's F score and the Index of Custodialism for his status membership is .46.

Relations between ideology and personality. One of our fundamental postulates is that an individual's ideological orientation is intimately bound up with his deeper-lying personality characteristics. We therefore hypothesize, at the collective level, relative congruence between modal ideology and modal personality. The obtained rank-order correlation between CMI mean (our measure of modal ideology) and F mean (our measure of modal personality) for the 12 units is .81. Thus, the congruence between modal personality and

[5] The use of indices based on rank in a product-moment correlation involves the assumption of equal intervals between ranks. This constitutes a possible source of error, but probably not a great one.

modal ideology in a system is relatively great. As noted earlier, the size of the variance on CMI is also associated with that on F.

With the individual hospital member as the focus of analysis, the CMI-F correlation for the total sample of 335 (regardless of specific status membership) is .71. We can now consider the relationship between ideology and personality when system membership is held constant. The F-CMI correlations for the single status groupings are presented in Table 2. They average .71, a value identical to the CMI-F correlations for the sample as a whole, and 11 of the 12 correlations are significant at the .05 level or better.

We thus have evidence that the differences in modal ideology among the 12 status units are closely related to differences in modal personality. When we find individual differences in ideology within a single status unit, these differences are closely related to differences in personality characteristics.

The theoretical formulations and results presented here concerning the mental hospital have their parallels in other social settings such as the school, the prison, industry, and the family. In all these institutions a small "administrative" elite has the power and responsibility to set goals and to control the destiny of a massive "membership." This larger population, whether children, patients, or prisoners, is a potential threat to society's values; various measures of education and social control are necessary. One of the major forms of conflict arising in these institutions is that between autocratic and democratic orientations. There is considerable evidence from both the present research and related studies that the autocratic-democratic ideological continuum is one aspect of a broader authoritarian-equalitarian personality continuum. Social ideologies have, to a considerable extent, a psychological basis in the personalities of their adherents. A sociopsychological approach provides, we believe, an important adjunct to historic-sociological approaches in the study of ideology.

SUMMARY AND CONCLUSIONS

This inquiry has taken as its starting point the distinction between "custodialism" and "humanism" in the mental hospital. These terms refer to two contrasting ideological orientations and to the corresponding forms of hospital policy. We have investigated ideology both as an individual and as a collective phenomenon—or, more accurately, we have used both individual and collective modes of analysis in the study of ideology. With regard to the individual, we have tried to assess ideology by means of a specially devised CMI (custodialism-humanism) scale, and to determine the relationships between ideology and other individual characteristics such as psychodynamics and membership in various groups. With regard to the collective unit (e.g., hospital or occupational status), we have tried to assess the degree of custodialism in its policy requirements and in the modal ideology of its personnel, as well as the degree of authoritarianism in the modal personality of its personnel, and to determine the relationships among these.

In the individual, preference for a custodialistic orientation is part of a broader pattern of personal authoritarianism. Correlations averaging about .70 were found between the Custodialism (CMI) scale and the scales measuring autocratic family ideology (TFI) and general authoritarianism (F). Although various hospital groupings differ significantly in mean CMI score, there are appreciable individual differences within most of the groupings studied. These ideological differences within single hospitals and occupations are quite closely related to differences in personality.

In the collective unit, we found relatively great congruence between prevailing policy, modal ideology, and modal personality. The hospital-status units having the most custodial policy requirements had as well the most custodial modal ideologies and the most authoritarian personalities. At the same time, it should be noted that the correspondence among policy, ideology, and personality is far from complete. Each of these aspects of collective life can vary to some extent independently of the others, and the phenomenon of incongruence is as important as that of congruence.

Although none of our groups can be regarded as ideologically homogeneous, some of them showed relatively small dispersion in CMI scores. These groups had a similar dispersion in F scores, and had low CMI and F means. Our data do not tell how the low diversity and the high ideology-personality congruence came about, but they point up

the need for answers to at least the following questions. To what extent do relatively homogeneous systems maintain themselves by recruitment and selective maintenance of individuals whose personalities are receptive to the structurally required ideology? To what extent do systems change the personalities which initially are unreceptive to the prevailing policy? Under what conditions can a system induce most of its members to support the required ideology even when this ideology is personality-incongruent? Under what conditions can the "incongruent" members change the system to a personally more congenial form?

Systems characterized by relatively great ideological diversity were very common in our sample. Moreover, the ideological diversity went hand in hand with diversity in personality, the standard deviations on CMI correlating .61 with those on F. We incline to the belief that significant heterogeneity of opinion and of personality obtains in the majority of institutional settings within modern societies undergoing rapid technological and educational change.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.
2. BETTELHEIM, B. *Love is not enough.* Glencoe, Ill.: Free Press, 1950.
3. CHRISTIE, R. Authoritarianism re-examined. In R. Christie & Marie Jahoda (Eds.), *Studies in the scope and method of "The authoritarian personality."* Glencoe, Ill.: Free Press, 1954.
4. CRONBACH, L. J., & MEEHL, P. E. Construct validity in psychological tests. *Psychol. Bull.*, 1955, **52**, 281–302.
5. DICKS, H. V. Personality traits and national socialist ideology. *Hum. Relat.*, 1950, **3**, 111–154.
6. FROMM, E. *Man for himself.* New York: Rinehart, 1947.
7. GILBERT, DORIS C. Ideologies concerning mental illness: A sociopsychological study of mental hospital personnel. Unpublished doctor's dissertation, Radcliffe Coll., 1954.
8. GREENBLATT, M., YORK, R., & BROWN, ESTHER L. *From custodial to therapeutic care in mental hospitals.* New York: Russell Sage Foundation, 1955.
9. JONES, M. *The therapeutic community.* New York: Basic Books, 1953.
10. LEIGHTON, A. H. *The governing of men.* Princeton: Princeton Univer. Press, 1944.
11. LEVINSON, D. J., & HUFFMAN, PHYLLIS E. Traditional family ideology and its relation to personality. *J. Pers.*, 1955, **23**, 251–273.
12. MAYO, E. *Human problems of industrial civilization.* Cambridge, Mass.: Harvard Univer. Press, 1933.
13. PINE, F. Conceptions of the mentally ill and the self: A study of psychiatric aides. Unpublished doctor's dissertation, Harvard Univer., 1955.
14. STANTON, A., & SCHWARTZ, M. *The mental hospital.* New York: Basic Books, 1954.
15. TANNENBAUM, F. *Crime and the community.* New York: Columbia Univer. Press, 1951.

Received December 1, 1955.

PERSONALITY STRUCTURE AND GROUP STRUCTURE: AN INTERPRETATIVE STUDY OF THEIR RELATIONSHIP THROUGH AN EVENT-STRUCTURE HYPOTHESIS[1]

ARNOLD S. TANNENBAUM
Survey Research Center, University of Michigan
AND FLOYD H. ALLPORT
Maxwell Graduate School, Syracuse University

THE RESEARCH to be reported in this article was designed, in part, to test the hypothesis that an individual's pattern of personality constitutes a basis from which one can predict his attitudes toward a work-program in which he is a participant. That there is such a relationship seems almost a matter of common sense. If the group situation is suited to the worker's personality, his attitude toward it, other things being equal, would be expected to be favorable; if it is unsuited, his attitude will probably be unfavorable. For such an hypothesis only confirming results would be anticipated.

Obvious as the above hypothesis may appear, however, the fact that some theoretical guidance is necessary for its testing is soon realized. The notion of "suitedness," for example, mentioned in the opening paragraph, is rather vague. In this problem we have the analysis of the collective work situation, the pattern of a worker's personality, and the degree, or intensity, of each of the characteristics that comprise that pattern. How can we proceed to derive from these items a formula that will show the degree of compatibility or suitedness between the collective situation and the worker's personality? Is suitability a quality, or can it be expressed, physiologically, in terms of energies? If suitedness means that the situation gives the worker "a chance for self-expression," how can self-expression be defined and measured? Then too, personality is a complex affair; it has many characteristics, each at its own level of intensity. How are these complexities to be accommodated and appraised in predicting the attitude? And once we have operationalized suitability, how shall we conceive the dynamics of attitude to be organismically related to it?

It is to questions such as these that the present article is addressed; and for a conceptual background and quantitative formulation that may be helpful in answering them the writers have made use of the theory of event-structure.

EVENT-STRUCTURE THEORY AND ITS RELATION TO ATTITUDES: LOGIC OF THE EXPERIMENT

A complete summary of the theory of event-structure will not be attempted here; we shall present only a few of the pertinent concepts.[2] A basic postulate of the system is that behavior can be described as a cycle of events or as a system of such cycles in interrelationship. The contention that the behavioral pattern is a *true cycle* is based on the fact (or likelihood) that the events which make it up return to the region in which they started. For example, "drinking" usually starts with drying in the mouth or throat, and the events that ensue (involving receptors, neurons, muscle fiber contractions, etc.) bring the cycle of ongoings back to the same region, as the membrane of the mouth and throat is moistened. This feature of return to the starting region is known as "closure." The throat-drying represents a "*primary*" event (really many events in a region), which constitutes the initial energic upset of the cycle. The act of drinking

[1] This study was conducted as part of a larger experiment directed by N. C. Morse and E. Reimer and constituting one of a series of researches, under the Human Relations Program of the Survey Research Center, on organizational functioning. The writers wish to thank Dr. Morse and Mr. Reimer for their generous help in the conduct of this study. They also wish to acknowledge a grant from the Rockefeller Foundation which contributed to its support.

The material in the present article was drawn from a dissertation submitted by A. S. Tannenbaum in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Graduate School of Syracuse University, February, 1954.

[2] For more complete details of event-structure theory, so far as it is at present formulated, the reader is referred to two publications by F. H. Allport (3, 4). The reader is also referred to a mimeographed document which discusses in greater detail the theory as it pertains to the present research (5).

involves a displacement or equalization of energy through the cycle which reduces the energy of the primary region. It is hypothesized that all behavioral acts, overt or covert, can be described in terms of such self-closed cycles or structures.

Event-structure theory, which is a general conception, provides for the "group" or collective situation in the same way. It is conceived as a structure of a higher order that is built, on the same principles, from the structures of individuals' behaviors. Events occur between individuals; and, depending on probabilities, the events proceed from energic upset in one part (or individual) around the collective cycle and back to the initial region, in which an intraorganismic equilibrium of the participant is then attained. We have, therefore, a *collective* structure (in our present case, the work group) as well as the individual behavior-structures that compose it.

Another important feature of the theory is its interpretation of meaning. It is hypothesized that a behavior cycle, with repetition, can come to be abridged in energies and in the number of events that involve contacts with the environment. It may be limited to the proprioceptive circuits, or even further restricted to the cerebrum or to the cortex alone. Such a cycle represents the "meaning" of the original overt act. It can also represent such acts in a "generalized" way. It still, however, is a true cycle, having a primary event region and a closure which occurs as the cycle completes itself and comes to equilibrium. Subjectively we say, at such a time, that a "meaning" is experienced.

An aspect of personality is just such a "meaning-cycle." It is one that is characteristic of the individual. We shall call such a personality-characteristic a *trend-cycle* or *trend-structure*. In more familiar language, a trend-cycle refers to something which the individual is "characteristically trying to do" (though not necessarily consciously trying), that is, to some meaning that he is continually seeking to attain or realize through his behavior (1). Its energy level indicates "*how hard*" he is trying to achieve that meaning. The individual, however, has *more than one* trend; and their cycles are conceived as forming a trend-system. In terms of the theory, it is this system which must be considered as basic in the influence of an individual's personality upon his attitudes toward the collective situation. One of the variables concerned in the predictive formula of our problem will be the energies which we refer to as the "potency" of the personality trends of the subject.

A trend-cycle usually achieves closure not by itself, but through some (overt) behavioral cycle which affords the "means," that is, through a pattern of events, by which the closure of the trend can be accomplished. Such an "instrumental" cycle or set of cycles may be provided by a collective structure. To illustrate this latter point, let us suppose that a certain player in a volley-ball game has a strong trend-cycle of "trying to cooperate." The collective structure of the game consists of a series of events involving close teamwork on the part of each player; and this fact gives our subject, as he operates in the structure, a continual opportunity for the closure of this trend. The trend cycle and the game-playing cycle are here said to be "interstructurant." We say, more familiarly, that the game affords an opportunity for "expressing this characteristic (or meaning) of his personality."

But there is still another feature to be considered. Not all overt or collective structures give the same opportunity, or means, for the closure of a given trend-cycle. We have therefore to consider the "*degree of interstructurance*" between the trend-cycle and the tangent collective structure.[3] We shall here refer to this aspect as the *relevance* (*R*) of the collective structure to the trend.

Finally we come to the question of attitudes, our dependent variable. An attitude is considered to be a reaction of favor or disfavor toward a social object or symbol, in this case toward the collective work structure. An attitude is energically bound up with both the individual's trend cycles and the collective structure in which he is involved. Let us suppose that a worker has a strong trend of "trying to rely on others." Now if the organization of the group work-program is such that it requires continual close supervision and direction by a group leader, the program will be highly relevant to this trend; and we might expect that there would be a considerable probability of the trend's closure. This circumstance, combined with the high potency of the trend-cycle, would be expected to add a con-

[3] Symbolized as "I" in the references cited.

siderable increment to the favorableness of the worker's attitude toward the program. But let us take the case of another worker who has a trend "to be treated by others as an equal." The types of work-setting just mentioned would probably yield little if any relevance to the trend, and the energic increment to attitude from this trend would be negligible, or perhaps even negative. Suppose that the nature of the program were such as to leave most decisions to the workers, with little or no supervision. Here the relevances would be reversed for the two individuals, and the predictions for attitude would be correspondingly altered.

The specific illustrations given above compare the attitudes of two different individuals. It must be remembered, however, that the same individual may have trend-cycles (of differing potencies) to which the work situation has different degrees of relevance; and the individual also has many trends. *Summation* of all these energic effects will therefore be required.

The task of predicting the attitudes of an individual toward a collective structure in which he is operating, then, must provide at least for the following procedures: (*a*) the obtaining of a set of measurements of his trend-potencies and of the degree of relevance of the collective structure to each of his trends; (*b*) the combining of these two measurements in some way for each trend; and (*c*) the combining of all the resulting quantities into some over-all formula by which the individual's attitude can be predicted. We shall be concerned first with the problems involved in the last two steps.

Derivation of a Predictive Formula

As stated earlier, the amount of closure of the trend, a factor upon which the related attitude variable is to be predicted, can be defined as the effect of the collective structure in reducing the primary-event energy of the trend and bringing that structure to equilibrium. Such a measure of the closure effect would be given by subtracting from the trend's initial energy (potency) the amount of energy left in the trend-structure after some standard period of performance of the individual in the collective structure. This quantity can be expressed by

$$C = P_1 - P_2, \qquad [1]$$

where C represents the amount of the trend's energy closed; P_1 the amount of the trend's energy at time 1, before closure; and P_2 the amount of its energy remaining after the collective structure's operation (i.e., after closure).

It is our assumption that attitudes can be understood in terms of the elements of this formula. However, we would predict that an attitude toward a collective structure would be a function, not only of the amount of closure afforded by the collective structure, but a negative function of the amount of trend energies remaining unclosed as a result of the operation of the collective structure. The favorableness of the attitude, then, would be predicted to be proportional to the amount of closure minus the amount of unclosure.[4] Using the symbol A_o to indicate the predicted energy of the attitude toward object o (the collective work situation), and remembering that we are dealing with the increment to the attitude given by only one trend, we have the formula

$$\text{Increment to } A_o = (P_1 - P_2) - P_2. \qquad [2]$$

Whenever the value of this equation is negative we might suppose that the direction of the attitude-increment is negative, that is, unfavorable to Object o. But the total energy of the attitude is made up of a summation of the increments from all the trends. Using a subscript T to represent each of the trends in the total personality structure, taken one at a time in the computation and summation, the formula then becomes

$$A_o = \sum (P_{1T} - P_{2T}) - \sum P_{2T} = \sum (P_{1T} - 2P_{2T}). \qquad [3]$$

And again, where $\sum (P_{1T} - P_{2T})$ (closure) is less than $\sum P_{2T}$ (energy remaining in the trend), thus making the right side of the equation negative, we might predict the attitude to be unfavorable to o.

Equation 3 deals only with the trend-potencies of the individual. We must now return to the task of bringing the relevance of the collective structure to these trends into the quantitative picture. In terms of the theory, both these factors help to determine the favorableness of an attitude. As will be presently shown, the relevance of the collective structure

[4] For an earlier application of this formula see N. C. Morse (8).

to a trend-structure was experimentally defined in the present study, and estimated as the probability that there are means available in the collective structure for the closure of the trend's energy. That is, if the probability of complete closure were rated as 1.0, it would mean that the rater felt certain that this energy would be fully closed—100 per cent of the trend's energy would be reduced by the collective structure. If the probability of complete closure were rated as 0, one would expect that even though some of its energy might be closed, the proportion would be minimal. Other probability ratings, between 0 and 1.0, could be taken as representing the likelihood of proportions of closure that lie between the minimal and the complete. If, then, R represents the *proportion* of the trend's energy (P_1) which becomes closed, the *amount* so closed becomes P_1R: the postulated relation is multiplicative. Using C for amount of closure, t for a particular trend-structure, and W for the collective work structure, we have the following formulation:

$$C = P_{1t}R_{W \to t}, \qquad [4]$$

or, in terms of the many trends which make up the total personality structure,

$$C = \sum P_{1T}R_{W \to T}, \qquad [5]$$

where C is total amount of closure and T, as before, is a class symbol for the trends whose respective PRs enter into the summation.

It will be remembered that the relation hypothesized between trend-energies, alone, and attitudes was

$$A_o = \sum (P_{1T} - P_{2T}) - \sum P_{2T}, \qquad [3]$$

and that closure, on a purely trend-energy basis, was defined as

$$C = (P_1 - P_2). \qquad [1]$$

For a number of trends the total amount of closure, on this basis, would be expressed by

$$C = \sum (P_{1T} - P_{2T}). \qquad [6]$$

From 5 and 6 we have the equality

$$\sum (P_{1T} - P_{2T}) = \sum P_{1T}R_{W \to T}. \qquad [7]$$

Substituting from 7 into 3

$$A_o = \sum P_{1T}R_{W \to T} - \sum P_{2T}. \qquad [8]$$

But since

$$\sum P_{2T} = \sum P_{1T} - \sum (P_{1T} - P_{2T}) \text{ (identity)},$$

or

$$\sum P_{2T} = \sum P_{1T} - \sum P_{1T}R_{W \to T},$$

Equation 8 now becomes

$$A_o = 2\sum P_{1T}R_{W \to T} - \sum P_{1T}. \qquad [9]$$

The right-hand side of this equation is the independent variable by which the attitude of the worker is to be predicted. The equation now states the structural hypothesis in terms of variables for which suitable measures can be found, viz., the potencies or energies of the subjects' trend-cycles, ratings by judges of the relevance of the collective work-structure to the trend-structures, and the attitude-scale positions of the subjects. The operations for measuring these variables will be discussed in the following section.[5]

METHOD

Work-Structures and Subjects

Two contrasting types of collective structures were available for the testing of this hypothesis. These were group situations approximating respectively the democratic and autocratic forms of organization. The investigation was conducted in one department of a large business organization, and the subjects (*S*s) were women who were employed in clerical work (7, 9). Each of the two experimental programs established within the department included about 200 persons. The programs were made to differ in the way in which authority was distributed within them. In the working groups of one of the programs, which was called the "autonomous" program, the employees (clerks) were given greater control than they had previously exercised over some of the work conditions. The responsibilities of each employee were also given greater emphasis. Authority previously held by the company officers was delegated to division managers who, in turn, delegated authority to the section heads and employees as a group. The clerks in this program participated in group meetings at which they discussed topics of their own choice. They made decisions regarding how the work would be distributed, length of the recess period, their right to leave the department during working hours, and similar matters. In the second experimental division, which was called the "hierarchical" program, the activities of the clerks were regulated more rigidly. Some scientific management principles were incorporated. Lectures by com-

[5] In references 3 and 4 of the bibliography another formula with different notation is proposed. It is based upon a more systematic assembling of structures and a slightly different concept of interstructurance (relevance).

pany experts, work standards, and incentives were included. Decisions were made primarily by upper-level company officers. These two programs, at the employee level, continued for about one year, a time span that will be referred to as the experimental period.

Trend-Structures and the Measurement of Their Potencies (P_{1t})

A standard sample of 26 trends was used for all *S*s. Several judges (*vide infra*) considered a range of "possible" trends that were suggested by a systematic canvass of the likelihood that the work-program (autonomous or hierarchical) would have some appreciable relevance to them. Since the programs were opposite in character, trends having high relevance with respect to one of them would usually have low relevance with respect to the other. The outcome for the trends thus selected (Table 1) will be seen to provide an equal number of trends having an $R > .5$ and an $R < .5$ with respect to each of the work-programs. It is not maintained that this method of selection, and the trend-list that resulted, assured a truly representative sample of trends for the individual *S*s, but only that certain gross errors of sampling were eliminated by giving high and low relevance an equal chance to play a part in the predictive score.

In the relevance-rating procedure, three judges were employed who were both psychologically trained and familiar with the work situations.[6] The method included the drawing up of lists of roles which the *S*s were considered as playing in their respective programs. Starting with general categories, a level of specificity was finally contemplated that dealt with such work-structure occurrences as "clerk receives help from supervisor," "clerk participates in a group meeting," "clerk reacts socially with another clerk" and so on. In deducing the trends that might be related to these event-patterns, the judges tried to answer the following question: "What is the nature of the trend meaning-cycle which would obtain closure through this specific series of events?" Additional criteria employed were (*a*) likelihood of the trend's occurrence within the type of population studied, (*b*) the applicability of the trend deduced to the early years of life (since it is believed that characteristics of personality are as a rule, long standing), and (*c*) freedom from ambiguity. Twenty-six trends were finally selected for measurement. They are presented in the left-hand side of Table 1.

A questionnaire was designed to measure the strengths of these trends in each of the *S*s. Five questions were employed for each trend, examples of which are as follows: For the trend "trying to be independent of control figures" (Item 7 in Table 1), one of the questions was: "How important is it for you to feel that you can run your life without depending upon people who are older and more experienced than you?" For the trend "trying to understand others' viewpoints and feelings," the following "agree" or "disagree" item was one of the five employed: "I don't care too much what other people think and how they feel about things. What they think is not my concern." In one of the five questions for the trend "trying to keep one's role at any given time distinct from other roles," the subjects were asked: "If you are discussing something serious with a friend, how much would you object to her bringing in some humorous comments?" For the trend "trying to adhere rigidly to specific rules or directions," one question asked "How much do you dislike it if directions on a recipe or sewing pattern are only partially laid out?"

Although some variety was introduced in the form of the questions, the majority were to be answered by checking on a 5-step scale that contained appropriate words indicating degrees of frequency, amount, etc., in which the *S* had the feeling or tendency in question. Care was taken to make the questions general in form and distinct from the company situation; and the questionnaire was administered at a time considerably earlier than the measuring of the dependent, attitudinal, variables. A potency score (P_{1t}) was computed for each trend for each respondent by taking the mean of her scores on the five questions designed to measure that trend. Test-retest reliabilities for the trend-scores in the population (over a one-year period) varied from .38 to .78, with an average correlation for all trends of .55.

The Measurement of Relevance ($R_{W \to t}$)

Ratings of the relevance of work-structure to trend-structure were made for each trend, and with respect to each of the two work-programs (autonomous and hierarchical). They were made by two of the judges who had participated in the selection of the trends to be rated. In this procedure the relevance-judge estimated the probability that complete closure for the trend would be given during the course of a day's work in the program concerned. The estimates were made on a numerical scale of 1 to 5, but were later translated in terms of a probability continuum ranging from zero to 1.0, zero meaning no probability of complete closure and 1.0 meaning certainty of complete closure. The mean of the judges' ratings for a given trend, with respect to a given program, was taken as the relevance-score of the program to that trend. These relevance-scores, expressed as probabilities, are shown in the right-hand side of Table 1. The judges made their ratings separately, but after having been associated on the study for some time during which some common understandings no doubt developed. The product-moment correlations between the judges' ratings were .83 for the relevance of the autonomous program, and .87 for the hierarchical. The relevance-ratings were, of course, given only upon the basis of a knowledge of the work-program in relation to the trend under consideration, and never with any particular *S* in mind.

Assigning "Suitability" Scores

An "attitude-prediction" score was computed for each individual by substituting the appropriate values in the right-hand side of the formula, $2\Sigma P_{1T} R_{W \to T} - \Sigma P_{1T}$ (Equation 9). For the *S*s in the autonomous program the relevance ratings of that program to the trends ($R_{I \to t}$, in Table 1) were used, and in the hierarchical program, the ratings $R_{II \to t}$ were used. The prediction scores, computed in accordance with the formula, may be considered to indicate the "suitability" of the individual's personality structure to her work-

[6] N. C. Morse, E. Reimer, and A. S. Tannenbaum.

TABLE 1
TRENDS USED IN THE EXPERIMENT, WITH RELEVANCES OF THE TWO WORK-STRUCTURES

Trend The subject is *characteristically trying* to:	R of Autonomous Program to Trend ($R_{I \rightarrow t}$)	R of Hierarchical Program to Trend ($R_{II \rightarrow t}$)
1. assume responsibility (anticipating, preparing for, and accepting the consequences of one's activities)	.84	.32
2. be creative (develop new patterns; to use one's imagination; to think up new and original ways of doing things)	.72	.00
3. increase the variety (number of types) of satisfactions in any given situation (to enjoy as many different things as possible in the same situation, e.g., to "mix business with pleasure")	.67	.19
4. understand and explain the reasons for one's behavior and opinions (to understand and explain one's feelings)	.82	.32
5. help others	.69	.32
6. understand others' viewpoints and feelings (to know why people feel or act as they do; to understand the reasons for others' behavior)	.82	.32
7. be independent of control figures (to be free of directions and control of persons in authority)	.82	.13
8. express one's considered judgments (to think through things carefully and state one's opinions)	.82	.19
9. base one's actions on one's own critical judgments and evaluations (to carefully evaluate and decide things for oneself)	.57	.07
10. act on the same level as others (to treat and be treated by others as an equal)	.88	.34
11. understand relations between and reasons for things (to understand the basis for things; to figure things out; to know "why")	.75	.13
12. do one's own thinking (not to be suggestable; to decide things for oneself; to draw one's own conclusions)	.75	.38
13. take the initiative with others (to start things; to organize things)	.94	.13
14. avoid emotional involvements (to avoid the expression and reception of emotional acceptance; to keep emotionally distant from others; to avoid being "warm" in relationships with others; to avoid affection)	.07	.63
15. conform to the wishes of control figures (to please persons in authority)	.13	1.00
16. adhere rigidly to specific rules or directions (to follow rules to the letter; to be given directions for doing things)	.32	1.00
17. act inferior to control figures (to be humble in relation to persons in authority)	.13	.64
18. avoid committing oneself (to keep one's opinions to oneself; to keep from getting oneself "out on a limb"; to keep people from knowing where one stands on controversial questions)	.25	.94
19. keep one's roles (behavior categories, relationships) at any given time distinct from other roles (to keep various parts of one's life separated)	.25	.88
20. admire or respect control figures or symbols (to look up to persons in, or symbols of, authority)	.19	1.00
21. be superior to others (to be better than others; to be more outstanding than others)	.25	.98
22. be efficient (not to waste time or effort)	.32	.88
23. be submissive to control figures (to submit to dominant leaders; to obey strong and forceful authority figures)	.13	.63
24. show self-discipline (to show self control; to keep one's behavior disciplined)	.50	.88
25. depend on others (to receive direction from others)	.38	.82
26. obey rules and follow directions	.25	1.00

program structure (I or II) as represented by W in the formula. Those *Ss* whose suitability scores were above the median of that variable's distribution in their program were called "suited," and were predicted to fall into the upper half of the attitude distribution (i.e., to have relatively favorable attitudes toward their program). Those who were below the median of the suitability distribution in their program were called "unsuited," and were predicted to fall into the lower half of the attitude distribution (i.e., to have relatively unfavorable attitudes). It should be noted that the attitudes here referred to were those which express favor or disfavor toward the work-program in which the individual is placed. The extent to which the *Ss*' attitudes toward their program, as independently measured, correspond to this prediction, constitutes the operational test of the hypothesis. More specifically stated, we are testing the hypothesis that a greater proportion of those relatively "suited" to the program in which they are placed will be more favorable (or less unfavorable) to their program than will be the case for those who are "unsuited" to their program.[7]

[7] The hypothesis is stated in *relative* terms since other factors also would be expected to influence the absolute level of the subject's attitude.

TABLE 2

DESIRED LENGTH OF PROGRAM IN RELATION TO PERSONALITY SUITABILITY

(Values represent percentage of *S*s in each group)

Program	Personality	Relatively Long	Relatively Short	*N*
Autonomous	Unsuited	36.7	63.3	100% ($N = 30$)
	Suited	76.0	24.0	100% ($N = 50$)
Hierarchical	Unsuited	23.1	76.9	100% ($N = 13$)
	Suited	33.3	66.7	100% ($N = 15$)
Combined Data	Unsuited	26.2	73.8	100% ($N = 43$)
	Suited	60.0	40.0	100% ($N = 65$)

The Dependent Variable: Attitude Structures and Their Measurement

The attitudinal measures were obtained by means of a paper-and-pencil questionnaire administered at the end of the experimental period, about one year subsequent to the securing of the personality measures. Some of the attitude-measures were relatively direct, in that the respondent was asked specifically about her liking for the program. Others were less direct, but still bore on the question of favor or disfavor toward the program. The specific items, together with the method of scoring, will be described in connection with the results. The attitude data shown below were derived from four questions, each of which was designed to measure an aspect of the central dependent variable. The results on these questions are, of course, not independent; they represent simply a number of ways of appraising the attitude-dimension with which we are concerned.

RESULTS

The results are given in the form of 2 × 2 tables. Three tables are presented for each attitude-item: one for the data from the autonomous program, one for the hierarchical program, and one for both programs combined.[8] Each of the distributions, for both the independent and the dependent variable, has been dichotomized at a point as close to the median as possible. The terms "suited" and "unsuited," and "relatively" (e.g., "relatively long" or "relatively short") are used in the tables to indicate the two halves of the distribution. This procedure means that the break in the dependent-variable distribution in one program may be at a point different from that in the other, or different from that in the combined data. Statistical tests for reliability of differences between proportions were computed (6, pp. 75–77).[9] Since the hypotheses are directional, one-tailed tests were used. Results below the 5 per cent level of confidence are referred to as significant, while results between the 5 per cent and 10 per cent levels are considered "suggestive."

Desired Length of the Program

Each respondent was asked, "How long would you like the program to last?" This was a "write-in" question coded on a 5-point scale from "end right now" (value of 1) to "indefinitely" or "forever" or "always" (value of 5). The results for this item are presented in Table 2.

A large and significant difference is found between suited and unsuited persons in the autonomous program. Seventy-six per cent of the suited individuals wanted the program to last a relatively long time, whereas only 36.7 per cent of the unsuited wanted it to last so long. The z value for this difference is 3.49. In the hierarchical program, although the difference is in the predicted direction, it is not significant.[10] In the combined data, however, it is found that a significantly greater proportion of the suited individuals (60.0 per cent) desire a relatively long extension of the program. The proportion for unsuited individuals is much smaller (26.2 per cent). The value of z is here 3.45.

Satisfaction with the Way the Program Operates

A second attitude-item asked "How satisfied are you with the way things are operating now in your division and section compared to the way they were operating two years ago?"

[8] Pooling the data is legitimate, since the two programs represented two different populations who were oriented toward two different attitude objects, and with no overlapping on any of the terms of the predictive equation.

[9] The formula employed was

$$z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}.$$

[10] The reasons for the small N's include the following: the fact that only those subjects were used who had spent a full year in the program and for whom a full series of questionnaires was obtained and matched, the occurrence of failures to answer write-in questions, and irrelevant responses that could not be coded.

TABLE 3
SATISFACTION WITH THE WAY PROGRAM OPERATES IN RELATION TO PERSONALITY SUITABILITY
(Values represent percentage of *Ss* in each group)

Program	Personality	Relatively Satisfied	Relatively Unsatisfied	*N*
Autonomous	Unsuited	31.8	68.2	100% ($N = 44$)
	Suited	42.2	57.8	100% ($N = 57$)
Hierarchical	Unsuited	53.9	46.1	100% ($N = 26$)
	Suited	63.3	36.7	100% ($N = 30$)
Combined Data	Unsuited	48.6	51.4	100% ($N = 70$)
	Suited	59.8	40.2	100% ($N = 87$)

Answers were checked on a 5-point scale ranging from "much less satisfied with the way things operate now," to "much more satisfied with the way things operate now." Table 3 presents the data for this item.

Although the direction of the difference between suited and unsuited is as expected, the result for neither program proves significant or suggestive. When the data for the programs are combined however, a suggestive relationship is found in the predicted direction, with 48.6 per cent of the unsuited and 59.8 per cent of the suited persons being relatively satisfied with the way their program operates now as compared with two years ago ($z = 1.40$).

Degree of Liking for the Program (First Form of Question)

As a further exploration of the *Ss*' attitudes the following free-answer question was asked: "How would you describe the present program in your division?" The write-in answers were categorized by coders in six groups, arranged on a scale indicating the degree of liking manifested, and ranging from "like it very much," or "it's very satisfactory," or "never better," to "dislike it very much," or "don't like it at all," or "it's very poor." The results are shown in Table 4.

In the autonomous program proportionately more suited individuals (60.0 per cent) appear to have a greater liking for the program than unsuited individuals (37.5 per cent). The z value for this difference is 1.41, a "suggestive" finding. The difference in the hierarchical program is not significant or suggestive, but when the data are combined we find that a significantly greater proportion (47.2 per cent) of the suited individuals have more liking for their program than was the case for unsuited (23.1 per cent). The z value is 1.94.

Degree of Liking for the Program (Second Form of Question)

An additional question was asked in order to ascertain the degree of liking for the program, this time in terms of the individual's "feeling" about it. The *S* was asked: How do you feel about it (the program) now?" This free-answer question was coded for liking manifested on a 5-point scale ranging from a response that

TABLE 4
LIKING FOR PROGRAM IN RELATION TO PERSONALITY SUITABILITY
(First form of question; values given are percentages)

Program	Personality	Relatively Like	Relatively Dislike	*N*
Autonomous	Unsuited	37.5	62.5	100% ($N = 16$)
	Suited	60.0	40.0	100% ($N = 25$)
Hierarchical	Unsuited	40.0	60.0	100% ($N = 10$)
	Suited	36.4	63.6	100% ($N = 11$)
Combined Data	Unsuited	23.1	76.9	100% ($N = 26$)
	Suited	47.2	52.8	100% ($N = 36$)

TABLE 5
LIKING FOR PROGRAM IN RELATION TO PERSONALITY SUITABILITY
(Second form of question; values given are percentages)

Program	Personality	Relatively Favorable	Relatively Unfavorable	*N*
Autonomous	Unsuited	22.6	77.4	100% ($N = 31$)
	Suited	31.8	68.2	100% ($N = 44$)
Hierarchical	Unsuited	20.0	80.0	100% ($N = 15$)
	Suited	45.5	54.5	100% ($N = 11$)
Combined data	Unsuited	65.2	34.8	100% ($N = 46$)
	Suited	80.0	20.0	100% ($N = 55$)

indicated "dislike it very much" to a response indicating "like it very much." The data are presented in Table 5.

Neither program yielded significant or suggestive differences between suited and unsuited. When the data are combined, however, a significant relationship in the predicted direction is found. Proportionately more suited individuals (80.0 per cent) feel more favorable toward the program they are in than do unsuited individuals (65.2 per cent). The z value is 1.67.

SUMMARY AND CONCLUSIONS

The attitude of favor or disfavor, developed by workers in a large business organization to two experimental programs with contrasting patterns of allocating authority appears to be a function, as was predicted, of the interaction between the personality structure of the individual and the structure of the work-program in which he is operating. Those individuals, who by their trend-structures (potencies and relevance both being considered) are "suited" to the program they are in, tend to feel a greater satisfaction with it and a greater degree of liking for it than do those whose trend-structures are "unsuited" to their program. Suited individuals also tend to want their program to last longer. Though these differences are not always large and are not typically significant in considering one program by itself, they are (with one exception) in the expected direction. And when the data for the programs are combined, yielding a larger N, the difference is significant for three of the attitude-measures and suggestive for the fourth. It might be noted, also, that these differences occur despite the fact that, because of the homogeneity of the populations, the initial variances in the personality measures were rather small.

In developing the hypothesis tested in this experiment we have made use of the theory of event-structure. It may have occurred to the reader, however, that the results can be interpreted equally well in terms of other current conceptions of behavior, notions which are much simpler than the theory here presented. It might be contended, for example, that the judges of relevance in the experiment were merely successfully appraising the way in which the worker perceives the group process in relation to the fulfillment of his motivating needs; and attitude in current theory has for some time been regarded as a function of motive-strength and perceived goal-relevance.

The theory, however, does not lose its significance because of the consideration that results in the same general direction could have been predicted by other, more familiar, conceptions. Every theory requires empirical support for its predictions, and two different theories might require, for confirmation, the same direction of results. The advantages of the event-structure approach, we believe, include the following: (*a*) It offers an explanation of *how* the interaction between personality and group-action might take place. (*b*) It provides a rationale for mathematical treatment. (*c*) It is capable of providing generalizations that are broader than attitude or motive, even broader than the field of behavior.

The present test of event-structure theory is limited in scope and other theories might have predicted the same direction of results. In these respects, therefore, the support given it by the data is not unique or conclusive. The results lie in the direction predicted by the theory; but further investigations will be required before the merit of the structural explanation can be properly appraised.

REFERENCES

1. ALLPORT, F. H. Teleonomic description in the study of personality. *Charact. Pers.*, 1937, **5**, 202–214.
2. ALLPORT, F. H. An event-system theory of collection action: with illustrations from economic and political phenomena and the production of war. *J. soc. Psychol.*, 1940, **11**, 417–445.
3. ALLPORT, F. H. The structuring of events: outline of a general theory with applications to psychology. *Psychol. Rev.*, 1954, **61**, 281–303.
4. ALLPORT, F. H. *Theories of perception and the concept of structure.* New York: Wiley, 1955.
5. ALLPORT, F. H. *Theoretical statement concerning the relationship between personality structure and group structure.* Maxwell Graduate School, Syracuse Univer., 1956 (mimeographed).
6. MCNEMAR, Q. *Psychological statistics.* New York: Wiley, 1949.
7. MORSE, NANCY C., & REIMER, E. The experimental change of a major organizational variable. *J. abnorm. soc. Psychol.*, 1956, **52**, 120–129.
8. MORSE, NANCY C. *Satisfaction in the white-collar job.* Ann Arbor: University of Michigan Press, 1953.
9. MORSE, NANCY C., REIMER, E., & TANNENBAUM, A. S. Regulation and control in hierarchical organizations. *J. soc. Issues*, 1951, **7**, 41–48.

Received November 29, 1955.

STIMULUS GENERALIZATION AS A FUNCTION OF CLINICAL ANXIETY[1]

GERALD ROSENBAUM

Wayne University

THE concept of anxiety has been widely employed by psychologists to explain a variety of behavioral phenomena, ranging from the development and maintenance of various symptomatologies to the learning of the conditioned avoidance response. The clinical notion of anxiety generally corresponds to that of an implicit state or process within the organism which is inferred or defined on the basis of the observation of certain manifest symptoms. On the other hand, the term anxiety has been used by Mowrer (9) and Miller (8) in theoretical interpretations of learning phenomena to mean an internal response conditioned to previously neutral cues which have been associated with noxious stimulation.

Recent animal experiments (1, 5, 8) have shown that the operations defining anxiety in learning theory studies produce a state which exhibits both the reinforcing and energizing properties of an acquired drive. Evidence has been obtained by Welch and Kubis (12) and by Taylor (11) with human *S*s indicating that clinically defined anxiety may also exhibit these motivational properties. If, as there is reason to believe, the operations involved in either definition result in functionally equivalent drive states, then predictions involving the assumed relationship of the drive variable to other psychological variables should be equally applicable to both experimentally induced and clinically defined anxiety.

It has been previously reported (10) that, following Hull's postulates concerning the theoretical relationship between drive and effective habit strength, increases in experimentally induced anxiety produce a general elevation in the stimulus-generalization gradients of response amplitude and frequency obtained from human *S*s. The effects of two levels of clinical anxiety upon gradients of generalization will be considered in the present paper.

METHOD

Experimental design. Seventy-two male psychiatric patients from the Veterans Administration Hospital at Coatesville, Pennsylvania, and 72 male students from the introductory psychology course at the State University of Iowa served as *S*s for the experiment. In the first part of the experiment, 36 patients were designated as the high-anxiety group and 36 patients as the low-anxiety group on the basis of psychiatric ratings. The 36 patients in each group were then randomly assigned to one of three experimental conditions: (*a*) strong shock, (*b*) weak shock, or (*c*) buzzer. Thus, there was a high-anxiety and a low-anxiety subgroup of 12 patients for each of these conditions, allowing an independent evaluation of the effects of the two levels of clinical anxiety, of the three experimental anxiety conditions, and of their interaction.

The second part of the experiment employed the 72 students in an exact replication of the above design. Anxiety level was determined, however, by means of a scale (11) which had previously been used to discriminate high and low anxiety in a normal college population. The replication served to determine whether clinical anxiety would have the same relative effects in both pathological and normal *S*s, as well as providing a basis for the comparison of results obtained by the use of conventional psychiatric ratings of anxiety with those obtained from a more objective definition of anxiety.

Subjects. The selection of *S*s with high clinical anxiety from among the hospitalized patients was accomplished by means of a simple scale for rating overt evidences of anxiety, developed in collaboration with the psychiatric staff doing the rating. Patients who showed no indications of manifest anxiety were matched for psychiatric diagnosis with individuals in the high-anxiety group so that differences obtained between these groups might not be attributable to pathological variations other than anxiety among these *S*s. Of the patients selected, 40 were diagnosed as schizophrenic, 22 as psychoneurotic, and 10 as psychopathic personalities. The mean age of the high-anxiety group was 30.2; that for the low-anxiety group was 34.4. The mean IQ's for the two groups were 108.8 and 104.0, respectively. Neither of these differences between the groups was statistically significant.

Clinical anxiety level in the college population was defined by means of a modified form of the Anxiety scale described by Taylor (11). Fifty of the MMPI items which had shown the highest correlation with total anxiety score on the original scale were used in the

[1] This study was part of a doctoral dissertation completed in 1950 in the Department of Psychology at the State University of Iowa, under the direction of Drs. J. S. Brown and I. E. Farber. The author regrets the delay in publication and wishes to acknowledge those investigators (2, 6, 13) who have utilized the unpublished manuscript as a point of departure for further studies.

TABLE 1

MEAN AMOUNTS OF GENERALIZATION OF RESPONSE AMPLITUDE

Group	Normal *Ss*		Psychiatric *Ss*	
	High anxiety	Low anxiety	High anxiety	Low anxiety
Strong shock	22.71	14.74	18.89	12.98
Weak shock	11.39	12.59	12.50	10.91
Buzzer	12.80	10.86	11.08	11.70
All conditions	15.63	12.73	14.16	11.86

modified scale. In addition, the *F*, *K*, and *L* scales of the MMPI were included to detect false high- or low-anxiety scores. Any one of the following criteria served to invalidate an anxiety score: (*a*) an *F* score of 12 or above, (*b*) a *K* score of 24 or above, or (*c*) an *L* score of 7 above. The new version of the test was administered to 574 introductory psychology students, and the obtained distribution of scores was almost identical to the distribution obtained by Taylor. Thirty-six male *Ss* from the upper 20 per cent of the distribution and 36 males from the lower 20 per cent comprised the high- and low-anxiety groups, respectively.

Procedure. The apparatus employed in the experiment and the details of the procedure have been described previously (10). In general, the procedure involved training *Ss* to make a simple motor reaction as quickly as possible upon the presentation of a rectangularly shaped visual figure. Three other rectangles, differing from the training stimulus in height (in steps of 0.25 inches), served as the generalized stimuli to which the *Ss* were instructed *not* to respond. Following 20 trials with the training stimulus, *Ss* were given five test trials on each of the four stimuli. These 20 test trials were interspersed among 60 additional training trials, so that each test trial was followed by three presentations of the training stimulus. The frequency, amplitude, and latency of "false" motor responses to the test stimuli were obtained by means of a movement-recording pulley system and served as the measures of generalization. Variations in the level of experimentally induced anxiety were produced by instructing *Ss* that slow responses would elicit either a strong shock, a weak shock, or the sound of a buzzer. Shocks and buzzes were presented during the instruction period and on the sixth and eighteenth training trials.

RESULTS

The major results of the experiment are presented in Tables 1 and 2, which show the mean amounts of generalization of response amplitude and the percentage frequencies for each of the six subgroups of normal and psychiatric *Ss*. Each of the means in Table 1 is based upon an over-all generalization score obtained for every *S* by summing the mean amplitude of his responses to the four stimuli on the 20 test trials; the means in Table 2 were similarly obtained from the total frequency of responses during the 20 test trials, converted to percentages. Thus, these sum scores computed for each *S* represent essentially the heights of the individual generalization gradients, and their means designate the heights of the subgroup gradients. Latency data are not presented, since generalization gradients comparable to those found with the two principal measures were not obtained.

Main effects of clinical and experimental anxiety. The principal hypothesis of the present study was that high levels of clinical anxiety, as well as experimentally induced increases in anxiety, tend to elevate gradients of generalization. The *Ss* in the high-anxiety subgroups should therefore show greater amounts of generalization than those in the low-anxiety groups. Examination of Tables 1 and 2 reveals that the high-anxiety *Ss* tended to show more generalization of amplitude and percentage frequency in both normal and psychiatric groups. Statistical analyses of these data, a sample of which is presented as Table 3, indicates further that the difference between the two levels of clinical anxiety in amplitude of response is significant at the .05 level in the normal groups.

The difference between the two clinical-

TABLE 2

MEAN AMOUNTS OF GENERALIZATION OF RESPONSE FREQUENCY IN PERCENTAGE

Group	Normal *Ss*		Psychiatric *Ss*	
	High anxiety	Low anxiety	High anxiety	Low anxiety
Strong shock	73.75	57.10	65.00	50.85
Weak shock	39.60	46.25	47.50	42.90
Buzzer	42.90	38.75	40.05	48.35
All conditions	52.08	47.37	50.83	47.37

TABLE 3

ANALYSIS OF VARIANCE OF AMPLITUDE DATA FOR NORMAL SUBJECTS

Source of Variation	*df*	Mean Square	*F*
Clinical anxiety	1	151.67	3.99*
Experimental anxiety	2	371.77	9.78**
Clinical × experimental	2	130.43	3.43*
Within groups	66	38.01	
Total	71		

* $p = <.05$
** $p = <.01$

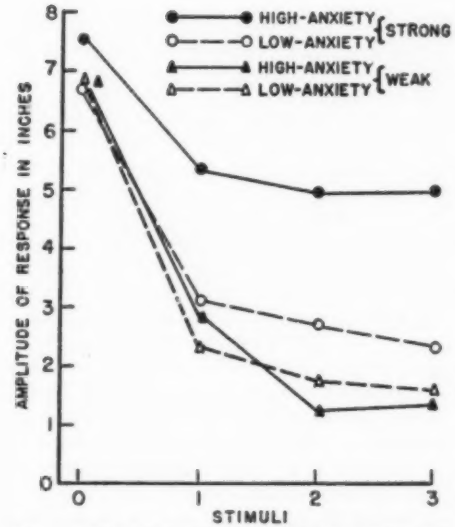

FIG. 1. GRADIENTS OF RESPONSE AMPLITUDE FOR HIGH- AND LOW-CLINICAL ANXIETY SUBGROUPS IN THE NORMAL GROUP

The training stimulus is designated as 0; and the generalized stimuli are designated 1, 2, 3, in order of decreasing similarity to the training stimulus. Weak-shock and buzzer conditions have been combined in the curves labeled weak.

anxiety levels in generalization of amplitude, however, was not significant in the psychiatric *S*s. Furthermore, the clinical anxiety variable produced no significant difference in mean frequency scores in either the normal or psychiatric *S*s. In contrast to these generally negative findings for the main effects of clinical anxiety, experimental anxiety produced significant increases in generalization of amplitude and frequency for both normal and psychiatric *S*s. As is apparent from Tables 1 and 2, however, while the strong-shock condition resulted in significant increases in generalization, the differences between the weak-shock and buzzer condition were not statistically significant.

Interaction of clinical and experimental anxiety. Further examination of Table 3 indicates that the interaction between clinical and experimental anxiety was significant at the .05 level. Careful scrutiny of Tables 1 and 2 and additional statistical comparisons reveal that this interaction effect is attributable to the fact that the high-clinical-anxiety groups exhibited significantly more generalization than the low-anxiety groups *only* under the strong-shock condition.

This effect is presented graphically in Fig. 1, which shows the gradients of response amplitude obtained from the normal *S*s under different experimental conditions. The weak-shock and buzzer groups have been combined in the curves labeled "Weak" for purposes of graphic presentation, since the differences between them were not significant. From Fig. 1 it is clear that strong shock consistently elevates the generalization gradient, but high clinical anxiety produced more generalization than low anxiety only when combined with strong shock. No significant difference in the height of the over-all generalization gradients was obtained between the high- and low-anxiety subgroups for either the weak-shock or buzzer conditions. Statistically significant interactions, essentially paralleling the results shown in Fig. 1, were obtained for frequency and amplitude measures in both the psychiatric and normal *S*s.

DISCUSSION

It has been demonstrated in the present study that while experimentally induced increases in anxiety act as a drive in elevating gradients of generalization, high levels of clinical anxiety exhibited drive properties only when combined with strong shock. These results suggest that clinical anxiety may act to increase generalization only when the stimulus situation involves the threat of strong punishment. This finding is only partly confirmatory of the hypothesis that manifest anxiety symptoms indicate the presence of a strong anxiety drive, since the effects of this additional drive might also be expected to increase generalization under conditions of weaker experimentally induced anxiety.

One possible interpretation of this inconsistency is that clinical anxiety symptoms, as defined either by psychiatric ratings or scores on an anxiety scale, do indicate the presence of a drive-like anxiety state, but that the motivational properties of this state are not activated unless the conditioned anxiety response is elicited by appropriate cues or by sufficiently noxious stimulation. Cameron has described the chronic anxiety reaction as "characterized by the presence of persistently heightened skeletal and visceral tension . . . which predispose the individual to give exaggerated and in-

appropriate responses on relatively slight provocation" (3, p. 249). This description parallels the hypothesis that anxiety-produced increases in drives should elevate the generalization gradients of frequency and amplitude, since elevation of frequency would produce inappropriate responses and elevated amplitude would produce exaggerated responses to stimuli that are normally ineffective in eliciting responses from less anxious people. Thus, the results obtained suggest that clinical anxiety, as measured in the present investigation, may be equated with other drives as a state-variable within the organism only under the specified conditions in which particular anxiety-producing cues are present in the situation.

An alternative interpretation of the fact that no difference was found between the two degrees of clinical anxiety under weak shock conditions is that the criteria employed for determining the levels of clinical anxiety were inadequate. This possibility hardly seems likely, however, in view of the difference obtained under the strong-shock condition. Furthermore, the fact that both the psychiatric ratings and the anxiety scale produced similar results indicates that this interpretation is insufficient. In fact, the similarity of the results obtained by the use of these two techniques for identifying clinical anxiety suggests that they tended to identify the same process and thus contributes to the construct validity of the anxiety scale as a measure of symptomatic anxiety reflecting an acquired drive.

It may be of additional interest to note that the present findings agree only in part with those of several recent studies. While Wenar (13) found that both clinical anxiety and stimulus intensity increased generalization to temporal stimuli, he obtained differences between his high- and low-anxiety groups under weak-shock and buzzer conditions, as well as under strong shock. Since either a shock or a buzzer was the signal to react on every trial, and since *S*s were always notified of their reaction time in Wenar's procedure, it would appear that these conditions are sufficiently anxiety arousing to elicit differences between clinical anxiety levels that were not obtained under the less threatening conditions of the present experiment.

A more serious discrepancy with the present findings and also with those of Wenar, however, is a suggestion by Eriksen (4) that his results indicate that high-anxious *S*s exhibit *less* generalization than low-anxious *S*s. This inference is based, in turn, on his findings that hysteric *S*s showed more generalization than psychasthenic *S*s, and that there was a significant negative correlation between his hysteria-minus-psychasthenia dimension and the manifest anxiety scale. Since Eriksen's suggestion is the direct opposite of the present findings, those of Wenar, and those reported by Hilgard, Jones, and Kaplan (7) that high-anxious *S*s give more conditioned eyelid responses to a generalized stimulus, it seems likely that the discrepancy suggested by his results may have been an artifact of the range of anxiety scores included in his experimental sample. It would also appear reasonable to note that the greater generalization shown by his hysteric *S*s may well have resulted from their repressive tendencies producing greater inattention and consequently more false responses. If such a relationship prevailed between general responsiveness and hysteria, the positive relationship between anxiety and generalization, demonstrated both for clinical ratings and the manifest anxiety scale in the present experiment, might well have been obscured and even reversed.

SUMMARY

The present study was concerned with the effects of anxiety upon stimulus generalization. It was hypothesized that the state inferred from manifest clinical symptoms of anxiety would show functionally similar motivational properties to the state of anxiety defined in terms of an implicit response that has been conditioned to situations involving noxious stimulation. It was predicted that both types of anxiety would exhibit the energizing properties of a drive and therefore elevate response gradients of generalization.

Seventy-two male psychiatric patients and an equal number of college students were divided into high-clinical-anxiety and low-clinical-anxiety subgroups by means of psychiatric ratings and a scale designed to measure anxiety in students. Variations in experimental anxiety were effected within each group by the use of two different intensities of shock and a buzzer. The *S*s were trained to make a motor

reaction to a visual stimulus and then tested on other stimuli differing from the training stimulus in height. The amplitude and frequency of responses to these generalized stimuli were employed as the measures of generalization.

The results showed that the groups designated as high in clinical anxiety showed significantly more generalization than the low-clinical anxiety groups under the strong-shock condition. No difference was found between the two levels of clinical anxiety for either the weak-shock or buzzer condition.

Markedly parallel results were obtained by the use of psychiatric ratings for differentiating the degree of clinical anxiety in the psychiatric patients and by the use of a psychometric scale with the normal *S*s.

It was concluded that the heightened responsiveness of anxious individuals to a wide range of stimuli may be subsumed under the more general effects of increased drive upon stimulus generalization, but that clinically defined anxiety is activated as a drive variable only when certain noxious cues are present.

REFERENCES

1. BROWN, J. S., & JACOBS, A. The role of fear in the motivation and acquisition of responses. *J. exp. Psychol.*, 1949, **39**, 747–759.
2. BUSS, A. H. Stimulus generalization as a function of clinical anxiety and direction of generalization. *J. abnorm. soc. Psychol.*, 1955, **50**, 271–273.
3. CAMERON, N. *The psychology of behavior disorders.* New York: Houghton Mifflin, 1947.
4. ERIKSEN, C. W. Some personality correlates of stimulus generalization under stress. *J. abnorm. soc. Psychol.*, 1954, **49**, 561–565.
5. FARBER, I. E. Response fixation under anxiety and non-anxiety conditions. *J. exp. Psychol.*, 1948, **38**, 111–131.
6. FARBER, I. E., & SPENCE, K. W. Relation of stimulus intensity, practice, anxiety, and sex to simple RT and temporal generalization. Iowa City: State Univer. of Iowa, June 27, 1955. (Tech. Rep. No. 4, Contract N9 ONR-93802.)
7. HILGARD, E. R., JONES, L. V., & KAPLAN, S. J. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, **42**, 94–99.
8. MILLER, N. E. Studies of fear as an acquirable drive: I. Fear as motivation and fear reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, **38**, 89–101.
9. MOWRER, O. H. A stimulus-response analysis of anxiety and its role as a reinforcing agent. *Psychol. Rev.*, 1939, **46**, 553–565.
10. ROSENBAUM, G. Stimulus generalization as a function of level of experimentally induced anxiety. *J. exp. Psychol.*, 1953, **45**, 35–43.
11. TAYLOR, JANET. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.
12. WELCH, L., & KUBIS, J. Conditioned PGR in states of pathological anxiety. *J. Psychol.*, 1947, **23**, 83–91.
13. WENAR, C. Reaction time as a function of manifest anxiety and stimulus intensity. *J. abnorm. soc. Psychol.*, 1954, **49**, 335–340.

Received November 18, 1955.

DISTRACTIBILITY IN THE CONCEPTUAL PERFORMANCE OF SCHIZOPHRENICS[1]

LOREN J. CHAPMAN

Northwestern University[2]

MANY studies have been carried out in which the performance of normal and schizophrenic *S*s on conceptual tasks have been compared. Despite a variety of techniques, nearly all such investigations have found that the schizophrenic group performs in an inferior manner to a normal group. Various writers disagree, however, on the cause of the inferior performance. One point of view, espoused by Goldstein and Scheerer (3), Hanfmann and Kasanin (4), and others is that schizophrenics are deficient in ability to form concepts. Other workers, among whom Cameron (1, 2) has been outstanding, believe that the schizophrenics do poorly, not because of inability to form concepts per se, but because of behavior tendencies which interfere with adequate performance. Cameron believes that schizophrenics are overly responsive to distractions provided by both external and internal stimuli, and in responding to such stimuli, produce responses which are incompatible with correct solution. Thus, a schizophrenic may be able to form the necessary concepts in a test, but fail because response to distracters prevents the demonstration of his concept-forming ability by a correct performance.

This latter interpretation suggests a possible experimental arrangement for providing evidence concerning the role of interfering factors in schizophrenic conceptual performance. If instead of requiring *S* to form a concept, *E* pointed out to *S* the correct concept and asked *S* to use it to guide his performance, schizophrenics might be expected to become increasingly inferior to normal *S*s in following these instructions as more and more distracting elements were introduced into the context of the situation. Assuming that the required correct solutions were the same throughout all degrees of distraction, such a differential rise in errors for the two groups could not be predicted from an inability of the schizophrenics to form concepts, since only the context of the task and not the task itself would be modified.

This distraction hypothesis could be tested in a number of ways, the simplest perhaps being variation of the number of distracters.

The dimension actually used in this study was a modification of this procedure, involving a card-sorting task in which *S* was required to assign several series of response cards (here called RCs), each bearing four figures, with one of three standard cards (SCs), each one also with four figures on it. Fig. 1 shows some examples of the cards. The *E* asked *S* to sort each RC on the basis of some commonality with an SC, but in doing so to use only the lower right-hand corner of the SC, ignoring the other three figures. (By making the correct sorting basis always the lower right-hand corner, *E* minimized the importance of memory in following the instructions.) Each RC contained one figure that enabled it to be correctly placed with one SC. Its other three figures, however, shared varying numbers of commonalities with figures on a second SC that were not in the lower right-hand corner and which, therefore, were incorrect for sorting. There were two kinds of RCs, namely identity RCs and concept RCs. On an identity RC, of which RC 24 in Fig. 1 is an example, one figure was identical to an SC figure that was correct for sorting (in the lower right-hand corner of the SC) and had other figures that were identical to figures on a second SC, these SC figures being incorrect for sorting (not in the lower right-hand corner of the SC). For example, RC 24 has an octopus which is identical to the octopus in the lower right-hand corner of SC I, and also has an apple and a hammer, identical to figures in positions incorrect for sorting on SC III. The letter *Z*

[1]This paper is based on a Ph.D. dissertation at Northwestern University. The manuscript in its present form was written while the author was a Public Health Service Research Fellow of the National Institute of Mental Health. The author is indebted to Dr. Janet A. Taylor for invaluable assistance and advice in directing the dissertation, and to both Dr. Taylor and Dr. Howard F. Hunt for their critical reading of the manuscript.

[2]Now at the University of Chicago.

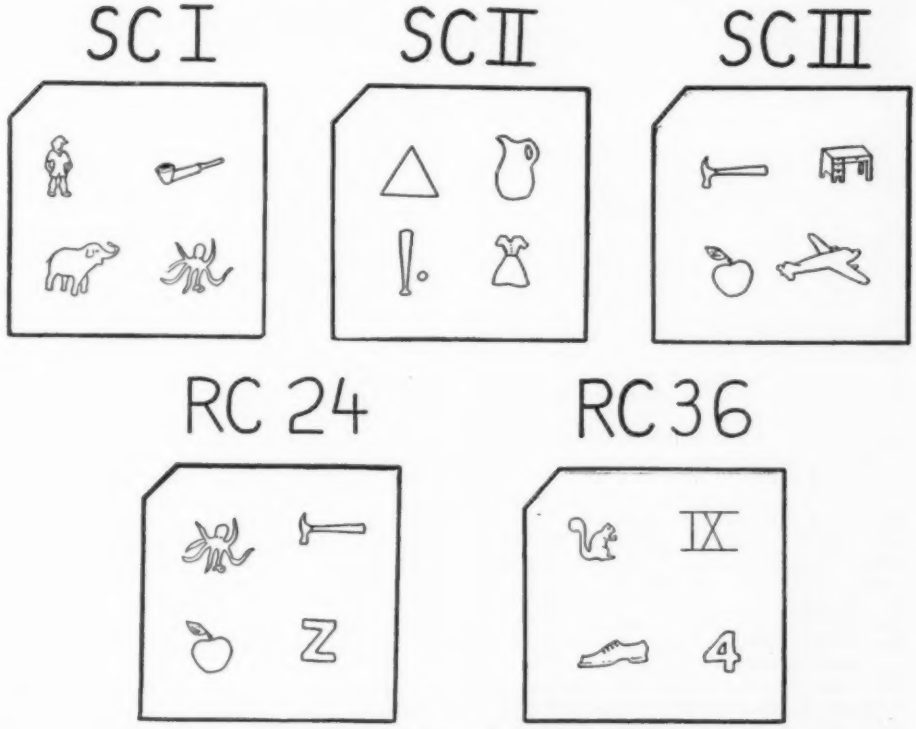

FIG. 1. EXAMPLES OF THE STANDARD CARDS AND RESPONSE CARDS

on RC 24 is a "filler" figure, designed to share no commonalities with SC figures but used on those RCs with less than three distracter figures to keep the number of such figures for each RC at four.

A concept RC, of which RC 36 is an example, differs from an identity RC only in that the commonalities both for correct sorting and for distracters are conceptual. For example, RC 36 has a shoe which shares a concept with the dress which is in the position correct for sorting on SC II, and also has a squirrel which shares a concept with the elephant, which is in a position incorrect for sorting on SC I. The Roman numeral IX and Arabic number 4 are filler figures on RC 36.

The number of commonalities present which were incorrect for sorting, here called the distracter level, was varied from 0 to 3. The prediction was that the schizophrenics would respond much more than normal *S*s to these distracting commonalities. No prediction could be made concerning a difference between the two groups at the zero distracter level because the task might be such a simple one (both in difficulty of concepts and in terms of distracters) that schizophrenics might do as well as normal *S*s. Rather, it was predicted that regardless of the comparative level of performance of the groups initially, the errors of the schizophrenics would rise with an increase in distracters while the errors of the normal *S*s would not, producing a divergence in the error curves.

One reason for making the choice of manipulating the number of commonalities rather than sheer number of figures was the greater meaningfulness that the use of this latter dimension would give to positive results. If it were found that schizophrenics but not normal *S*s make increasing errors with an increase in the number of distracters involving conceptual commonalities, it would give further evidence that the schizophrenic de-

fect in conceptual performance is not wholly one of concept forming; before a response to conceptual distracters can deter performance, *S* must obviously formulate a concept.

METHOD

Subjects. The final experimental groups, after pretesting, consisted of 25 normal and 40 schizophrenic *S*s. The normal *S*s were 21 Naval Corps School students at Great Lakes Naval Training Station and 4 Downey VA Hospital aides.

Schizophrenic *S*s were chosen from those patients at the Downey Hospital bearing the diagnosis of chronic schizophrenia whose case material did not indicate either cerebral organic involvement or mental deficiency. Other prerequisites were that the first hospitalization for psychosis had occurred at least one year previously, that the most recent hospitalization had occurred at least 3 months previously, that the patient had not received electric shock therapy within the past 3 months and had never received more than 35 treatments. No selection was attempted on the basis of diagnostic subtype.

All *S*s were males between 18 and 51 years of age and had between 7 and 16 years of education. The mean age, education, and functioning IQ for the schizophrenic group were 29.6, 11.4, and 95.1. The corresponding values for the normal group were 23.7, 10.6, and 99.9. The IQ scores were obtained from the CVS for the schizophrenic *S*s and hospital aides, and from GCT equivalents for the 21 naval personnel. Results of *t* tests and *F* tests showed that the groups did not differ significantly on mean or variability in either education or IQ. On age, the schizophenics were found to have a higher mean and a lower variability. However, there seems to be no reason to suspect that the age variable is correlated with response to the experimental variable being tested, so that the discrepancy between groups was not considered to be important.[2]

Experimental task. The task materials consisted of a number of cards, each 4½ inches square with four figures drawn on it in India ink. There was a total of 60 different figures used on the cards, the figures being chosen so as to fall into 15 conceptual categories of four figures each. With the help of a preliminary study, the conceptual categories and their member figures were chosen with a view toward keeping the individual figures recognizable and the concepts relatively nonoverlapping and simple.

Of the 15 categories, 12 were used as sorting categories, i.e., used as concepts according to which *S*s were expected to sort. These sorting categories were smoking equipment, clothing, geometric figures, sports equipment, furniture, means of transportation, tools, fruit, animals, sea life, people, and kitchen utensils. The other three categories, used as "filler" figures and never employed as sorting concepts but placed on some of the response cards to keep constant the number of figures at four, consisted of letters, Arabic numerals, and Roman numerals.

For each sorting task, a set of three SCs was used. For all sortings, the same 12 figures were used on the SCs, each of the figures being from a different category. Four different sets of SCs were obtained by moving the four figures on each of the three cards one position clockwise. One of the four figures on each SC, always the lower right-hand corner figure, was arbitrarily designated as the "correct" basis of sorting a particular series of RCs.

In addition to a correct and a distracter SC for each identity or concept RC, there was a third SC containing no figures that shared identities or concepts with it. This was called the irrelevant SC, and provided an index of errors other than those resulting from sorting with distracter figures. Which one of the three SCs fulfilled each of these three relationships varied from RC to RC.

Apparatus. Three 5½″ × 8½″ wooden boxes were placed on a table in front of *S*. On the vertical face of each box was a slot in which the RCs were to be deposited. On the top of each box there was a rack for holding the SC.

Preliminary training and screening. Before the experiment proper, two preliminary tasks were given both to pretrain *S*s in the materials used and to identify those *S*s unable to handle them. The *S* was shown the 60 figures used in the experimental task, each on a separate card, and asked their names. If *S* failed to give an approximately correct name he was told the answer. However, if *S* failed more than five figures on his first try (not including values of Roman numerals), or any one figure on a second presentation, he was dropped from the experiment None of the normal group and six of the schizophrenic *S*s were eliminated on this basis.

Next, *E* gave *S* a preliminary sorting task, again using the 60 single figure cards. The *E* placed four stimulus cards from different categories on the table and asked *S* to take cards from the top of the pile in order and place them with that stimulus card which pictured something of the same kind. When *S* had placed three additional cards with any stimulus card, *E* removed this category and replaced it with a new stimulus card that started a new category. The order of the cards in the pile was such that *S* was always able to place each card with one of the four choices confronting him. If necessary, *S* was helped or corrected, but if this happened on more than five cards, he was excluded from the experiment. None of the normal *S*s, but five schizophrenic *S*s were eliminated on this basis.

Presentation of experimental task. The *S* was instructed to choose the correct box for each RC by finding a figure in one of the four corners of the RC that was either identical to or of the same sort as the figure

[2] Matching of groups on any characteristic presupposes, of course, that the characteristic is correlated with the variable under experimentation. In the present study, since the experimental dimension was a new one, the only choice was to match on such obvious (but not necessarily related) factors as IQ and education. The discrepancy in age was not deemed relevant. A check on the data of this and a later experiment by the writer revealed no relation between response to distracters and age for either the normal or schizophrenic *S*s. In fact, for both groups of the present study, the younger *S*s actually made a few more errors, though the difference was not significant. Greater errors on the part of the younger *S*s would, if present, reduce the likelihood of obtaining the predicted difference between groups.

in the lower right-hand corner of one of the three SCs. The *E* asked *S* to explain these instructions before he began sorting. Only those *S*s were used who verbalized that they were to use the lower right-hand corner of one of the SCs and any corner of the RC. None of the normal *S*s, but 19 of the schizophrenic *S*s were dropped on this basis. The *S* was again asked to explain the instructions after his first series, and also after his last. On one or the other of these two repetitions, 8 of the schizophrenics failed, but they were not excluded from the experiment since they had all adequately explained the instructions on the first request before beginning to sort.

Each *S* received one 36-card series from each of the four distracter levels, each such series being divided into a subseries of 18 concept RCs and a subseries of 18 identity RCs.

Order of presentation of the series representing the four distracter levels was counterbalanced among the *S*s. Appropriate counterbalancing was also used to prevent biasing of results by preference for any of the box positions, figure position on the RCs, any of the various concepts and figures, or any differential difficulty in using the above, or by practice and fatigue effects.

RESULTS

Since the data were extremely skewed, nonparametric statistical techniques were used in all analyses.

Four types of error scores were used. The number of incorrect placements of the RC with the SC containing the distracter figures, Distracter Score, is a crude measure of sorting in response to distracters. The number of placements with the Irrelevant SC, Irrelevant Score, measures errors which are not in response to the distracter figures. These errors include mistakes due to carelessness, noncooperation, and sorting by commonalities which *S* may have perceived but which *E* did not intend to build into the test. Because, in any given subseries, errors of the irrelevant type should be made equally often in response to each of the three SCs, the difference between the two error scores, Distracter-minus-Irrelevant Score, is considered a purer measure than Distracter Score alone of sorting with the distracter figures. The final measure, Total Errors, is a measure of errors of all types. It is the only error score which can be used at the zero distracter level because without distracters there is obviously no Distracter Score.

Most of the analyses were done separately for the concept RCs (in which both the commonalities correct for sorting and the distracter commonalities were concepts), and for the identity RCs (in which both the correct and the distracter commonalities were identities).

TABLE 1

MEAN DISTRACTER AND IRRELEVANT SCORES FOR SCHIZOPHRENIC AND NORMAL *S*s ON CONCEPT AND IDENTITY RESPONSE CARDS (RCs) AT EACH DISTRACTER LEVEL

Distracter Level		Distracter Score		Irrelevant Score	
		Schiz.	Normal	Schiz.	Normal
0	Concept RCs			.98	.20
	Identity RCs			.72	.12
1	Concept RCs	2.15	.40	.75	.20
	Identity RCs	2.00	.48	.48	.16
2	Concept RCs	3.65	.68	.80	.08
	Identity RCs	3.87	.44	.48	.00
3	Concept RCs	3.22	.68	.55	.00
	Identity RCs	2.60	.40	.42	.12

Performance of the normal Ss. The normal group as a whole made very few errors. For the total of 144 card placements which each *S* made in the experimental task, 21 of the 25 *S*s made 2 or fewer errors. The remaining 4 *S*s erred, respectively, 3, 5, 7, and 65 times. The distribution of the Distracter and Irrelevant errors among the distracter levels for both concept and identity RCs is shown in Table 1, and the relationship between distracter level and Total Errors is shown in Fig. 2. Although there is a hint of an increase in Total Errors

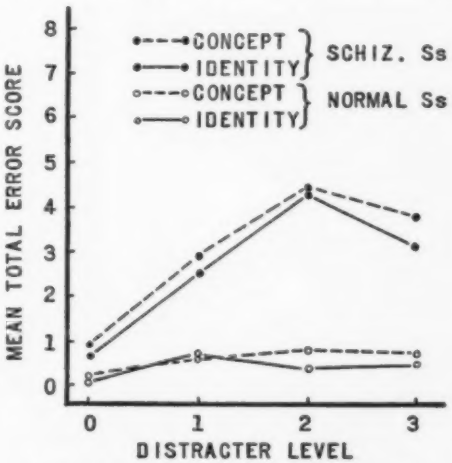

FIG. 2. RELATION OF TOTAL (T) ERROR SCORE TO DISTRACTER LEVEL

with the distracter level for these normal *S*s, due primarily to the single deviate normal *S* who made 65 errors, analysis of variance by ranks (5) showed for each of the two types of RCs no over-all difference on Total Errors among distracter levels. A similar lack of relationship was found between the other error measures and distracter level.

Performance of the schizophrenic Ss. Of the 40 patients, 18 made 3 or fewer errors, while the remaining 22 ranged as high as 108 errors in their 144 card-placements. As seen in Fig. 1, there is a strong relation between Total Error Score and distracter level, with the largest error difference occurring between distracter levels 0 and 1. Analyses of variance by ranks (5) showed an over-all significant difference between the four levels, both for concept RCs ($p < .001$) and for identity RCs ($p < .01$). In both cases, the rise was significant by a Wilcoxon test from 0 to 1 and from 1 to 2, but level 3 did not differ significantly from either 1 or 2. Thus, in general, the hypothesis is confirmed that as the number of distracters increases, the number of errors increases for schizophrenics but not for normal *S*s.

For the Distracter and Distracter-minus-Irrelevant Scores, inspection of Table 1 suggests a similar tendency for errors to rise from level 1 to 2 and a decrease, smaller in size, from 2 to 3. An analysis of variance by ranks, however, did not show an over-all difference among the three distracter levels. This lack of significance is not too surprising when one considers that a large part of the rise in errors with distracters is, as shown in Fig. 2, between the 0 and 1 distracter levels. Since the Distracter Score cannot be obtained at the 0-distracter level, the rise from 0 to 1 must be omitted from analysis on these two error measures, thus reducing the probability of obtaining over-all differences.

The data were examined to determine whether Total Error Scores increase within each 18-card subseries as a function of time since instructions, a finding which would be expected if the response to distracters resulted from a sheer forgetting of the instructions. The position of the card in the order of presentation was used as an approximate measure of time. When mean number of errors for each successive block of three cards was plotted, there was no progressive increase in errors within a subseries. This observation, together with the fact that only 8 of the *S*s failed to explain adequately the instructions on either the second or third request, indicates that forgetting of the instructions is of little or no importance in accounting for the schizophrenics' response to the distracters.

It will be remembered that many of the schizophrenic *S*s performed the experimental tasks with very few errors, equalling the performance of the normal *S*s in this respect. A further analysis of the schizophrenics' data was made to determine whether one of the factors contributing to response to distracters by a schizophrenic *S* was severity of illness. The patients were therefore divided into two groups, 23 patients given full privileges and 17 locked-ward patients. This division was used as a rough measure of severity of illness. They were also divided into 18 high scorers on Total Errors (12 or more errors on concept and identity RCs combined), and 22 low, and into 17 high Distracter - minus - Irrelevant scorers (score of 4 or more on concept and identity RCs combined), and 23 low. This latter grouping is of special interest here because it is based on errors independent of any difficulty which *S* may have had in handling the materials conceptually. Chi square showed that the locked-ward patients included more of the high Total Error scorers ($p < .05$) and of the high Distractor-minus-Irrelevant scorers ($p < .02$). Thus, even though the present experiment included only relatively undisturbed patients (the most disturbed patients in the hospital could not meet criteria for inclusion), as degree of schizophrenic pathology increased, so did response to distracters.

Comparison of normal and schizophrenic Ss. As has already been discussed, the performance of the schizophrenics was adversely affected by an increase in the number of distracters, while the performance of the normal *S*s remained constant. Turning to a more direct comparison of the performance of the two groups, the schizophrenics exceeded the normal *S*s on Total Error Score for the four distracter levels combined both on concept RCs ($p < .01$) and on identity RCs ($p < .01$), significance being tested by a Mann-Whitney *U* test. More enlightening is the comparison of the two groups at the separate distracter

levels. As seen in Fig. 1, at the zero distracter level, the schizophrenics made only a few more errors than the normal *S*s, and the difference fell short of significance both on concept RCs ($p = .44$), and on identity RCs ($p = .46$). At the 1-, 2-, and 3-distracter levels, however, the differences were much greater, and in each case, for the concept RCs, were significant at the .01 level or less. Thus, the presence or absence of distracters seems to be an important determining factor in whether or not the schizophrenics do more poorly than normal *S*s on a conceptual task.

On the Distracter Score, and the Distracter-minus-Irrelevant Score, the schizophrenics also exceeded the normal *S*s for the three levels combined (the zero-distracter level being necessarily omitted), both on concept and identity RCs ($p < .02$ or less), though for some of the levels taken separately the p values fell short of the .05 level of significance. The significant difference on the over-all Distracter-minus-Irrelevant Score for concept RCs is of special importance. This score measures errors in performance independent of errors arising from sources other than response to distracters; patients unable to grasp concepts, therefore, would be expected by chance to have Distracter-minus-Irrelevant Scores clustering around zero. Consequently, the greater error scores of the patients on this measure are clearly due to response to the distracters rather than to a lowered ability to sort in accordance with concepts.

DISCUSSION

The finding that the schizophrenics' errors rose as distracter level increased while those for normals did not confirms the hypothesis that much of the inferior conceptual performance of schizophrenics may arise from a response to distracters. This response may itself entail conceptual behavior as is borne out by the fact that the schizophrenics were affected adversely by the conceptual type of distracter as well as by the identity type. However, these results do not show whether or not there is also present in schizophrenics some lowered ability to form concepts. Not only do patients labeled schizophrenic differ a great deal among themselves in the kind of aberrant solution offered on conceptual tests, but individual schizophrenics often show a variety of such solutions. One must be cautious in elaborating any single finding to explain all such aberrations. Future research should indicate the importance of response to various kinds of distracters, as well as reveal other sources of schizophrenics' inferior conceptual performance.

SUMMARY

The hypothesis tested was that the inferior performance of schizophrenics on conceptual tasks is at least in part due to their handling the test materials in other ways than those required by instructions, rather than due only to an inability to form concepts. Normal and schizophrenic *S*s were asked to sort response cards containing four figures with three four-figured standard cards, one figure on each standard card being arbitrarily designated by instructions as correct for sorting with the rest incorrect. Each response card contained one figure that shared some commonality with the correct figure on one standard card. Also, the figures on each response card were varied so that from zero to three of them were identical to or shared concepts with incorrect figures on a second standard card. The schizophrenic group much more than the normal *S*s used the incorrect "distracter" commonalities as a basis for sorting, even when this required conceptual sorting. The error curves for the two groups therefore diverged as the number of distracter commonalities increased. The results are interpreted as supporting the hypothesis.

REFERENCES

1. CAMERON, N. Reasoning, regression and communication in schizophrenia. *Psychol. Monogr.*, 1938, **50**, No. 1 (Whole No. 221).
2. CAMERON, N. Deterioration and regression in schizophrenic thinking. *J. abnorm. soc. Psychol.*, 1939, **34**, 265–270.
3. GOLDSTEIN, K., & SCHEERER, M. Abstract and concrete behavior: an experimental study with special tests. *Psychol. Monogr.*, 1941, **53**, No. 2 (Whole No. 239).
4. HANFMANN, EUGENIA, & KASANIN, J. S. Conceptual thinking in schizophrenia. *Nerv. ment. Dis. Monogr. Ser.*, 1942, No. 67.
5. WALKER, HELEN M., & LEV, J. *Statistical inference.* New York: Holt, 1953.

Received December 8, 1955.

SCALING NORM CONFLICTS IN THE AREA OF PREJUDICE AND DISCRIMINATION[1]

NATHAN KOGAN AND JAMES F. DOWNEY
Harvard University

SINCE the publication of *The American Soldier* series, there has been a growing interest in the problem of multiple reference groups operating at cross purposes. Merton and Kitt (6) have been particularly forceful in calling for additional theory and research in this area, while others (10, 11) have designed ingenious questionnaires for tapping important aspects of norm and role conflict. Essentially, these investigators have sought to determine the manner in which discrepancies are resolved between divergent or contradictory norms prescribed by an individual's reference groups in a particular situation.

As of the present time, the norm-conflict model has not been employed in the context of prejudice and discrimination. Yet events on the American social scene suggest that such an application may yield dividends. In particular, incipient desegregation in public schools located in various parts of the South may result in conflicts between potentially tolerant attitudes toward Negroes and norms of intolerance fostered by parents, peers, and others in the community. Where there are attitudinal and behavioral changes toward minority group members in the direction of greater tolerance (or intolerance), the extent to which such change represents the resolution of a norm conflict could well bear investigation.

The present paper attempts the development of a scale measuring the propensities of individuals to choose a discriminatory, "hands-off," or militantly antidiscriminatory course of action in particular social situations. In the sense that the social situations are incidents depicting various kinds of prejudice and discrimination by parents, friends, and others, the choice of a course of action necessarily involves a distinctive resolution of a norm conflict.

METHOD

Subjects. The *S*s were recruited from the high-school population of a New England city. For many years, this city experienced considerable intergroup conflict. Segregation of ethnic and religious groups resulted in few intergroup contacts and considerable intergroup hostility. Occasionally, adolescent "gangs" organized along racial, religious, or ethnic lines erupted into overt "warfare" with each other. The prevalent norms about minority subgroups were of a distinct ethnocentric variety.

With the support of certain community leaders, a young female social worker established a youth club (approximately two years before the date of this study) with the intention of harmoniously bringing together adolescents in the various community subgroups. By introducing athletic, social, and community participation programs, she has succeeded in lessening intergroup tensions, and, to a real extent, has established a group norm of antiprejudice militancy.

At the time the study was conducted, the total membership of this youth organization exceeded 1,400 boys and girls. A sample of 100 (50 boys and 50 girls) ranging in age from 12 to 20 was selected from (*a*) the club leadership, (*b*) members who appeared at club headquarters after school hours, and (*c*) members who appeared at a weekend social function. Except for the exclusion of Negro members, the sample is fairly representative of the club population. A range of *S*s extending from leaders to persons on the periphery of the group is included in the sample.

The social situations questionnaire. The *S*s filled out a questionnaire in which they were required to choose one of four alternative responses for resolving fifteen anti-Negro discrimination situations. These conflict situations include various kinds of discrimination ranging from opposition to intimate social equality, at the one extreme, to the threat of physical violence, at the other. The situations are classified into three clusters of five items each in terms of the person or persons manifesting the discriminatory behavior. One item from each of these three clusters—peers, authorities, and strangers—is presented here in the order cited.[2]

> Imagine that you are planning to go out to a show with several friends and their dates. The day before you go out, one of the fellows calls you up and asks you if it would be all right if *S* (a Negro whom you do not know) could come along with his

[1] The authors wish to thank Miss Basilla Neilan for her cooperation in making subjects available for this study.

[2] The complete questionnaire, item-by-item response frequencies, and derived Guttman scales have been deposited with the American Documentation Institute. Order Document No. 4885 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C. remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

date. You're pretty sure that the other kids in the party would not like this. Under these conditions,

____I would definitely say "No" or else get out of the party.

____I would probably say I would rather not, because of how the other kids would feel about it.

____I would probably say that it would be OK with me.

____I would definitely say that it would be OK and would try to convince the others that this was a good idea.

Imagine that as you are sitting at home one day, a neighbor comes in to ask your parents to sign a petition which would prevent Negroes from buying or renting land on your block. He explains that it would not hurt the Negroes because there are plenty of other good places in town to live. This move would not only save your section just for White people, but would also make your land worth more money. Your folks are just about to sign the petition. Under these conditions,

____I would insist that they were wrong and try to persuade them not to sign the petition.

____I would probably tell my parents that I didn't think they were doing the right thing.

____I would probably keep quiet because it wouldn't make much difference to me one way or the other.

____I would definitely approve.

Imagine that you are visiting a friend of yours out of town. You decide to go swimming at a local pool. Getting into line to pay your admission, you notice a young colored boy in line just in front of you. As he gets up to the booth to buy his ticket, the woman in the booth tells him that they don't allow Negroes in the pool. Under these conditions,

____I would complain to the woman about this policy.

____I would voice a complaint to my friend making sure that the woman would hear.

____I wouldn't say anything about it then, but might make a criticism to my friend later.

____I would say nothing, feeling it is the right of the pool owners to reject Negroes if they wish.

Questionnaire administration and scoring. The *Ss* specified their age, sex, religious preference, and extent of participation in the youth group, but did not sign their names to the questionnaire. The social worker in charge of the youth organization was not present during the test administration and the respondents were informed that she would not see the questionnaires. Item responses were scored by means of Guttman scaling procedures (3).

RESULTS

Guttman scales[2] were obtained for the three item clusters and for all of the items taken together. The latter proved to be an 8-item scale with a coefficient of reproducibility of .91. The peers, authorities, and strangers items yield coefficients of reproducibility of .93, .96, and .95, respectively. All of the items in the peers and strangers clusters and all but one in the authorities cluster are represented in the subscales.

TABLE 1

PRODUCT-MOMENT INTERCORRELATIONS AMONG TOTAL SCALE AND SUBSCALES

	Total	Peers	Author-ities	Strangers
Total	—	.77	.51	.52
Peers	.82	—	.47	.44
Authorities	.55	.52	—	.37
Strangers	.55	.48	.41	—

Note.—The correlation coefficients above the diagonal are uncorrected. Those below the diagonal have been corrected for the small number of categories by means of the Peters and VanVoorhis (8) formula. All correlations are significant beyond the .01 level.

Despite high coefficients of reproducibility for the subscales, four or five dichotomized items are too few to guarantee unidimensionality. Guttman (4) maintains that the probability of obtaining a scale by chance is negligible for a sample of 100 on six dichotomized items, yet he questions the wisdom of generalizing this finding to an hypothetical item universe. While a reasonable degree of confidence attaches to the general scalability of norm conflict situations in the area of Negro-White relations, the positive findings respecting scalability within specific reference group clusters are necessarily of a tentative nature.

The intercorrelations among the scales are shown in Table 1. Differences in the extent of the correlation between the total scale and the subscales reflect the number of items contributed by each of the subscales to the total

[2] In accordance with common scaling procedure, unambiguous imperfect scale patterns were assigned to perfect scale types in such a way as to minimize error. Where imperfect scale patterns were ambiguous, i.e., could be assigned to more than one perfect scale type, the respondents were assigned to the scale type with the largest number of cases in accordance with Henry's (5) recommendations. Only those items meeting the "criterion of improvement" were included in the final scales.

In order to yield satisfactory Guttman scales, the four response categories had to be dichotomized as either 1 vs. 2, 3, 4 or 1, 2 vs. 3, 4 where 1 represents the most and 4 the least antidiscriminatory response. Very few *Ss* checked alternative 4, indicating that attitudes for the group as a whole are skewed toward the antidiscrimination pole.

scale. The peers subscale is the greatest contributor. The intercorrelations among the subscales are moderately high, suggesting that tendencies in the direction of discrimination or antidiscrimination cut across cluster areas.

To test the validity of the total scale, *S*s were placed in one of four activity groups depending upon the extent of their *actual* participation in the youth club. Activity group assignments were made before the determination of members' scale types. Group I included the leaders, while Group IV consisted of members who did not actively participate in any of the group's activities. Groups II and III represented intermediate levels of participation. The product-moment r between activity group and total scale type is .36, a value statistically significant at the .01 level. When corrected for the limited number of categories by the Peters and VanVoorhis (8) formula, r becomes .40. Product-moment r's between subscale types and activity group are .41 for peers, .44 for authorities, and .32 for strangers. Applying the same correction, the correlations become .47, .51, and .37, respectively. For each scale, *S*s assigned to the more militantly antiprejudiced scale types tend to be the more active members of the group.

DISCUSSION

With the demonstrated scalability of the norm conflict situations, one can begin to ascertain just how far an individual will go in resisting discrimination stemming from a variety of sources. Scalability of the item clusters, when firmly established, should permit an even more refined estimate by ordering the "conflictful" quality of situations within specific reference groups. In the peers subscale, for example, a progressively increasing number of *S*s endorses antidiscriminatory responses as item content extends from a disparaging joking remark about Negroes to the threat of physical violence.

The significant correlation between activity group and scale type supports the validity of the outlined instrument. The more active the group member, the stronger appears to be the incorporation of the youth-group norm.

The items developed in the present study are clearly most appropriate for adolescents who have some social contact with Negroes. However, where social class, geographical area, or other factors produce conflicts differing from those embodied in the items of the present study, it would be advisable to devise new items and scale them accordingly. As a methodological contribution, the present investigation makes no claims respecting generalizability to other subject populations or to different sets of items, but merely purports to demonstrate the potential unidimensional scalability of norm-conflict situations in the intergroup relations area.

It should be noted that the whole gamut of attitudes and behaviors subsumed under the rubric of prejudice and discrimination can hardly be forced into a norm-conflict mold. As Allport (1) has indicated, prejudice can be conceived as a continuum extending from sheer conformity to social norms on the one extreme to a high degree of functional significance for the personality at the other pole. While the present authors strongly endorse this view, they would maintain that conformity pressures may assume heightened significance in an adolescent group. For many of the members of the youth club described above, ethnocentric loyalties fostered by parents and peers were transformed into personal loyalty to a strong group leader (the social worker) with militantly tolerant attitudes. White adolescents in other communities may, to a considerable extent, be caught between existent conformity pressures toward discrimination and increasing tolerance induced by widening social contact in school and out-of-school situations.

Others (7, 9) have devised "social situations" tests for measuring behavioral predispositions in a variety of areas. Items for these instruments were selected by the Thurstone equal-appearing intervals method. The intention of those investigators was one of bringing attitude scales a step closer to the behavioral level. Somewhat earlier, Bogardus (2) devised his Social Distance Scale, an instrument with latent unidimensional properties. The present scale attempts to embody the advantages of these earlier instruments, while incorporating recent advances in unidimensional scaling techniques and the theory of social norms and reference groups.

There are distinct limitations in the use of a social situations instrument. The direct quality of the items suggests that some *S*s may respond in a deliberately self-enhancing manner,

whether this be in the direction of one or another set of norms. This problem, of course, is common to all attitude scales. Unlike attitude scales, however, the present instrument is explicitly intended as an indicator of prospective behavior. There is a distinct possibility that the checked alternative, though subjectively honest, is self-deceptive, for the *S*'s behavior in an actual situation might not coincide with his response on paper. Finally, a general activity vs. passivity or interventionist vs. noninterventionist dimension may account for some of the variance in the scores. Some *S*s would militantly follow the youth club's interests in any situation where the club clashed with persons or institutions external to it. The intercorrelations among the subscales are not so high, however, as to indicate that this militancy is an undifferentiated dimension.

Since the purpose of the present paper is essentially methodological, many substantive questions remain unanswered. While the potential unidimensional scalability of norm-conflict situations in the intergroup conflict area has been demonstrated, the reasons why specific situations are resolved in a specific direction for specific individuals have not been treated. Future theory and research, as Merton and Kitt (6) stress, should move on to consider "the dynamics of selection of reference groups among the individual's several membership groups."

SUMMARY

A sample of 100 high-school students—members of a youth club with a militantly antiprejudiced norm—filled out a questionnaire requiring them to choose one of four courses of action for resolving a series of Negro-White conflict situations. The questionnaire yielded an 8-item Guttman scale with a coefficient of reproducibility of .91. When the items were classified in one of three clusters in terms of the protagonist of the discriminatory act—peers, authorities, and strangers—these clusters also yielded Guttman scales with high coefficients of reproducibility. The subscale findings were rendered tentative by the relatively small number of dichotomized items in each of the three clusters. Intercorrelations between the total scale and the subscales and among subscales were moderately high. Scale validity is supported by a significant correlation between scale type and extent of participation in the youth group. A note of caution is sounded respecting generalizability of the present findings to other subject populations and item universes. The present scale is compared with relevant antecedent instruments and the scale's limitations are discussed.

REFERENCES

1. ALLPORT, G. W. Prejudice: A problem in psychological and social causation. *J. soc. Issues, Suppl. Series*, 1950, No. 4.
2. BOGARDUS, E. S. *Immigration and race attitudes*. Boston: Heath, 1928.
3. GUTTMAN, L. The Cornell technique for scale and intensity analysis. *Educ. psychol. Measmt*, 1947, **7**, 247–280.
4. GUTTMAN, L. The basis for scalogram analysis. In S. A. Stouffer, *et al.*, *Measurement and prediction*. Princeton: Princeton Univer. Press, 1950. Pp. 60–90.
5. HENRY, A. F. A method of classifying nonscale response patterns in a Guttman scale. *Publ. Opin. Quart.*, 1952, **16**, 94–106.
6. MERTON, R. K., & KITT, ALICE S. Contributions to the theory of reference group behavior. In R. K. Merton & P. F. Lazarsfeld (Eds.), *Studies in the scope and method of "The American soldier."* Glencoe, Ill.: Free Press, 1950.
7. PACE, C. R. A situations test to measure social-political-economic attitudes. *J. soc. Psychol.*, 1939, **10**, 331–344.
8. PETERS, C. C., & VANVOORHIS, W. R. *Statistical procedures and their mathematical bases*. New York: McGraw-Hill, 1940.
9. ROSANDER, A. C. An attitude scale based upon behavior situations. *J. soc. Psychol.*, 1937, **8**, 3–15.
10. STOUFFER, S. A. An analysis of conflicting social norms. *Amer. sociol. Rev.*, 1949, **14**, 707–717.
11. STOUFFER, S. A., & TOBY, J. Role conflict and personality. *Amer. J. Sociol.*, 1951, **56**, 395–406.

Received December 12, 1955.

DEVELOPMENT AND EVALUATION OF A SCALE FOR MEASURING SOCIAL ACQUIESCENCE[1]

BERNARD M. BASS

Louisiana State University

In a preceding report (3), the suggestion was made that acquiescence, defined as the tendency to accept any generalization about behavior, is a parsimonious concept that accounts for a large percentage of the variance among many personality measurements. Acquiescence to a variety of other stimuli has also been observed (1, 3, 7, 8, 9).

The forms of acquiescent behavior tend to be fairly independent of each other (11, 14). The purpose of the present study was to develop and evaluate a measure of social acquiescence —acquiescence to a wide variety of generalizations concerning how persons behave or should behave. In agreement with others (2, 5, 6, 7, 9, 11, 12), it was believed that developing such a measure should provide a useful personality assessment device for the social psychologist.

Method

Initial Subjects and Pool of Items

An initial sample of 200 college students, half from the South and half from the Midwest, indicated whether they agreed with each of 300 statements. Fifty-five per cent of the sample was male. The mean age of the subjects (*Ss*) was 21.0 years (SD = 3.5); the mean education was 14.1 (SD = 1.5).[2]

The *Ss* were directed to "indicate whether you agree, disagree or are uncertain about" each of 300 famous sayings. The statements (maxims, aphorisms, adages, proverbs, apothegms, and sayings) concerned thirteen areas of human behavior. Twenty statements involved each of the following: Material comfort, sex, harm avoidance, achievement, affiliation, deference, autonomy, aggression, abasement, rejection, nurturance, superego strength, and irritability. Forty additional statements, many opposed in meaning to each other, covered a wide miscellany of matters.

While it was shown elsewhere (4) that performance on the 13 scales could be accounted for by three common factors, in the present study, it was assumed that one general or second-order factor permeated all the scales as well as the 40 statements that had remained unclassified, the tendency to accept or reject the widely varying statements regardless of their content.

Item Analysis

The proportion of all 300 proverbs accepted by an *S* was used as a gross measure of the criterion against which a brief acquiescence scale was developed by item analysis. The 200 *Ss* were ordered according to these criterion scores. The performance of the upper 25 per cent of *Ss* was then contrasted with the performance of the lower 25 per cent on each of the 300 statements.

Fifty-six items emerged which were accepted by at least 40 per cent more of the upper than the lower criterion groups. These 56 statements, which are listed in Table 1, constitute the Social Acquiescence scale. The number of statements accepted among the 56 constitutes an *S*'s Social Acquiescence score. "?" and "Disagree" responses are ignored.

Results and Discussion

Reliability

A new sample of 50 West Coast residents and 50 Southern college students was selected with a mean age of 28.2 years (SD = 2.2). Forty-nine per cent were male. In this sample, the corrected split-half reliability of the 56-item Social Acquiescence scale was .92. A reliability of .81 was found for a more homogeneous sample of 1491 Louisiana college freshmen, using the Kuder-Richardson formula 21. The 56-item scale correlated .95 with tendency to accept or reject the 300 items.

Acquiescence or Acquiescences?

Consistent with the proposition that acquiescence may appear in a variety of independent forms, a correlation of .00 was found for 38 Louisiana State University night-school students between the Social Acquiescence scale and *Ss*' tendency to accept, agree, or react positively to 12 nonexistent test items which the experimenter supposedly had in

[1] These analyses were aided by grants from the Louisiana Graduate Council on Research. The author wishes to thank Kim Ki Suk, George Palmer, and Austin Flint who carried out the statistical analyses. He is also indebted to Donald Campbell, Gerald McCullough, Herbert Rothschild, Arnold Gebel, Perry Davis, Cecil Gibb, and Sam Hicks for their assistance in data collection.

[2] Heterogeneous samples were selected purposefully for the various analyses upon which the scale development was based. It was assumed that differences in one or more of personal characteristics such as age, sex, education, and regional background would yield increased personality variance. This increase would permit maximum discrimination for analysis and test validation. Moreover, the scale would be developed using samples representing the population for which the scale was being constructed.

TABLE 1
THE 56-ITEM SCALE OF SOCIAL ACQUIESCENCE

Success against odds is the greatest of American ideals.
Love is the greatest of Arts.
There is no satisfaction without a companion to share it.
Love of the opposite sex makes the world go round.
They never fail who die in a great cause.
He that has many friends need never fear disaster.
Destroyers of tyranny have contributed the most to mankind.
You only injure yourself when you take notice of despised critics.
The only known cure for fear is faith.
Our chief want in life is somebody who will make us do what we can.
Never trust a flatterer.
He who laughs last laughs longest.
No principle is more noble or holy than that of true obedience.
There is nothing which the body suffers which the soul may not profit by.
One false friend can do more harm than one hundred enemies.
No gift is more precious than good advice.
Obedience is the mother of success.
The victory always remains with those who admire rather than with those who criticize.
The greatest fortunes are for those who leave the common path and blaze a new trail for themselves.
'Tis vain to quarrel with our destiny.
To be happy, always stay within the law.
You should give more than you want to give.
Pity is the touch of God in human hearts.
What we win through authority we lose; what we win through consideration we keep.
A sense of duty is the basis of character.
Next to love sympathy is the most divine passion of the human heart.
Stay away from the proud man who is ashamed to weep.
Sweet is the sleep of men with virtue.
One should feel the failures of his friends as if the failures were his own.
Giving is always better than receiving.
He that loses his conscience has nothing left that is worth keeping.
Virtue is a struggle in which we overcome our weaknesses.
He conquers all who conquers himself.
It is difficult to do excellent work without great strain.
Only a statue's feelings are not easily hurt.
Happiness must be won through great effort.
The restless sleeper blames the couch.
Seeing is believing.
Still water runs deep.
Make yourself honey and the flies will eat you.
The grass is always greener in the other fellow's yard.
Most big cows have little horns.
Every man is blind to his own defects.
Jaws are the only part of the body that like to work.
Those in high places are in greater danger than those in lowly ones.
Life is a struggle from beginning to end.
Wild colts make good horses.
Empty heads go with loud talk.
You can't teach an old dog new tricks.
Count your sheep and the wolf will eat them.
Sleep is loved by everyone.
The feeling of friendship is like that of being comfortably filled with roast beef.
Who does not love the opposite sex remains a fool the whole life long.
Better one safe way than a hundred on which you are not sure.
We like best that which lies beyond our reach.
Amusement is the medicine for worry.

mind. Since both measurements were reliable, they appear to be consistent but independent measures of different forms of acquiescent behavior.

Acquiescence and the Antidemocratic Personality

Gaier and Bass completed an unpublished correlational analysis of the performance of 108 Maryland, 82 Kansas, and 71 Louisiana college students on the Social Acquiescence scale, the ethnocentrism or E scale and the authoritarianism or F scale. For the Maryland, Kansas, and Louisiana samples respectively, the F scale correlated with the Social Acquiescence scale: .49, .48, and .16. The E scale correlated with the Social Acquiescence scale .33, .20, and −.06. The E-F correlation for the three samples varied in the same way: .62, .60, and .49. Judgment is reserved concerning sampling effects on the intercorrelations. It is important to note here, however, that much E-F covariance both within and between samples could be accounted for by the mutual correlation of the two measures with Social Acquiescence. The Kansas sample means were highest while the Maryland sample means were lowest on all three scales.

Relations with Other Measures

Inspection of Table 2 and Table 3 suggests a pattern of relations between social acquiescence and miscellaneous aptitude, personality and behavior measures. One might infer that the person high in social acquiescence is an "outward-oriented," insensitive, nonintellectual, socially uncritical individual; in short, a Babbitt—an unquestioning conformer to social demands placed on him. While in many cases samples are too small to provide statistical significance, the findings will be reported briefly as suggestive for future investigation.

Ability and education. Correlations listed in Table 2 between various ability measures and Social Acquiescence tend to be low and negative. Inspection of the group differences shown in Table 3 suggests a fairly strong negative relation between amount of education and Social Acquiescence. College-trained *S*s exhibit less acquiescence than those who have not gone beyond high school.[3]

[3] The same negative relationship with educational level was found for the F scale in several studies. (See Cohn, T. S. and Carsch, H. Administration of the F scale to a sample of Germans. *J. abnorm. soc. Psychol.* 1954, **49**, 471.) The fact that the F scale and the Social Acquiescence scale are both negatively associated with education conforms to our hypothesis (3) that the F scale is primarily a measure of acquiescence, not authoritarianism.

TABLE 2

PRODUCT-MOMENT CORRELATIONS BETWEEN THE SOCIAL ACQUIESCENCE SCALE AND MEASURES OF ABILITY, SOCIABILITY, AND SOCIAL SENSITIVITY

Sample	Other Measure	Trait Assessed	Correlation with Social Acquiescence
27 industrial supervisors	Otis Gamma	intelligence	−.15
100 college students	ACE	intelligence	−.12
43 grocery salesmen	Wonderlic	intelligence	.00
100 college students	Grade-point average	academic achievement	−.12
77 factory supervisors	Supervisory aptitute battery	supervisory aptitude	−.23
77 factory supervisors	Years of education	education	−.12
41 night school students	Gordon Personal Profile-A	ascendency	.22
25 college students	GZ Temperament Survey-A	ascendency	.06
41 night school students	Gordon Personal Profile S	sociability	.34
25 college students	GZ Temperament Survey-S	sociability	.27
77 factory supervisors	Leadership questionnaire	favor initiative	.21
25 college students	Leaderless discussion success	initiative	.32
25 college students	GZ Temperament Survey-P	cooperativeness	.25
23 enlisted Marines	Peer nominations	"likes to help"	.45
23 enlisted Marines	Peer nominations	"thinks well of most"	.42
23 enlisted Marines	Peer nominations	"influences others"	.23
23 enlisted Marines	Peer nominations	"rugged individualist"	−.29
23 enlisted Marines	Peer nominations	"chip on shoulder"	−.24
25 college students	Kerr Empathy Test	empathy	−.14
25 college students	Accuracy on "How Others Rate Me"	empathy	−.34
25 college students	GZ Temperament Survey-T	introversion	−.25

TABLE 3

MEAN DIFFERENCES IN SOCIAL ACQUIESCENCE AMONG SELECTED SAMPLES

Sample	Region	Mean Acquiescence	SD
Adults			
43 Wholesale grocery salesmen	South	38.6	8.51
96 Penitentiary psychopaths	South	34.6	8.99
51 Penitentiary inmates without psychopathy	South	33.8	7.82
77 Industrial supervisors	South	33.7	9.18
18 Wholesale cosmetic salesmen	Midwest	33.6	8.33
34 Marine enlisted men	U. S.	31.9	10.40
49 Public school teachers	South	31.3	9.33
49 Residents	West Coast	31.2	11.24
Adolescents			
78 Rural high school seniors	South	38.3	9.36
145 Rural high school sophomores	South	36.7	8.97
36 Student nurses	New England	35.2	9.06
1,491 College freshman	South	32.6	8.09
100 College sophomores	South	32.1	11.60
100 College sophomores	Midwest	29.6	11.00
45 College seniors and graduate students	South	26.9	10.49

Personality. Correlations listed in Table 2 suggest that the socially acquiescent individual tends to be more sociable and yet more insensitive socially.

Population differences. The belief that the acquiescence scale is a measure of "Babbittry" is supported by a further examination of some of the differences among the groups listed in Table 3. Southern salesmen, much more likely to have Fromm's "marketable personality," show more Social Acquiescence than any other Southern group. Similarly, the small Northern group of cosmetics salesmen appear to be fairly high in Social Acquiescence for a non-Southern adult sample. Prisoners tend to be less acquiescent than nonprisoners matched for educational level (13). The greater tendency to acquiesce (about one third of a standard deviation) of Southern compared with Midwestern sophomore college students is consistent with their differences in background. In the South, there remains a more ready acceptance of the traditional and conventional manifested in the greater religiosity among Southern students as a whole, "official optimism" about social issues, and lack of skepticism.

One last analysis that supports the "Babbittry" hypothesis involved examination of the scores of 45 advanced college students who were asked to take the Social Acquiescence test twice, once imagining themselves applying for a job where they wish to appear as de-

sirable as possible, and once answering the questions as honestly as possible. The mean "faked" score of 30.3 was significantly greater at the .01 level than the mean "honest" score of 26.9, supporting the inference that high Social Acquiescence is regarded as desirable just as sociability, ascendency, emotional stability, and other such traits are usually thought by the layman to be related positively to career success.

Research Significance

The evidence presented on Social Acquiescence suggests the need to reinterpret much published social psychometric data. It is probable that Social Acquiescence accounts for much, if not all, of the covariance found between a wide variety of measures ranging from favoring birth control to belief in the efficacy of Yoga or conversely, disapproving desegregation to rejecting the use of American military force in Outer Mongolia.

There are at least four ways of eliminating or reducing the constant error due to acquiescent tendencies. First, new social psychological measurements can be constructed to balance carefully affirmative items with their opposites. Second, the reading ease of items can be reduced. Third, more intelligent and critical subjects can be employed. Fourth, acquiescence can be measured and then controlled by means of partial correlation or by sampling procedures.

SUMMARY

A 56-item scale of social acquiescence has been developed by internal consistency item analysis. It has a reliability above .90 and correlates .95 with the tendency to accept or reject 300 proverbs, aphorisms, and similar statements. According to population differences and correlations between the scale and various personality and intelligence measures, one who earns a high Social Acquiescence score resembles Sinclair Lewis' "Babbitt"—an unquestioning conformer to social demands. The Babbitt syndrome still must be regarded as a tentative sketch of the socially acquiescent *S* since many of the correlations reported were based on small samples and lacked statistical significance.

The demonstrated existence of consistent individual differences in tendency to acquiesce to generalizations about behavior has methodological importance. Previous research results based on the correlation between attitude scales (e.g., E and F scales) in which all statements are scored in the same direction, require reinterpretation.

REFERENCES

1. ANCONA, L. Indagine sulta natura psichica del "Response Set." La "motivazione al successo." *Arch. psicol. neurol. Psichiat.* 1954, **15**, 23–74.
2. BARNES, E. H. The relationship of biased test responses to psychopathology. Unpublished doctor's dissertation. Northwestern Univer., 1954.
3. BASS, B. M. Authoritarianism or acquiescence? *J. abnorm. soc. Psychol.*, 1955, **51**, 616–623.
4. BASS, B. M. Development of a structured disguised personality inventory. *J. appl. Psychol.*, in press.
5. BERG, I. A. Response bias and personality: the deviation hypothesis. *J. Psychol.*, 1955, **40**, 61–72.
6. BERG, I. A., & COLLIER, J. S. Personality and group differences in extreme response sets. *Educ. psychol. Measmt*, 1953, **13**, 164–169.
7. BERG, I., & RAPAPORT, G. M. Response bias in an unstructured questionnaire. *J. Psychol.*, 1954, **38**, 475–481.
8. COHN, T. S. Factors related to scores on the F (Predisposition to Fascism) Scale. Unpublished doctor's dissertation. Univer. of Michigan, 1953.
9. CRONBACH, L. Further evidence on response sets and test designs. *Educ. psychol. Measmt*, 1950, **10**, 3–31.
10. FLEISHMAN, E. The measurement of leadership attitudes in industry. *J. appl. Psychol.*, 1953, **37**, 153–158.
11. GAIER, E. L., LEE, M. E., & MCQUITTY, L. L. Response patterns in a test of logical inference. *Educ. psychol. Measmt*, 1953, **13**, 550–567.
12. LORGE, I. Gen-like: Halo or reality? *Psychol. Bull.*, 1937, **34**, 545–546.
13. PALMER, G. Discrimination of psychopaths, normal prisoners and non-prisoners using a disguised objective personality test. Unpublished master's thesis, Louisiana State Univer., 1956.
14. SINGER, W. B., & YOUNG, P. T. Studies in affective reaction: III. The specificity of affective reactions. *J. gen. Psychol.*, 1941, **24**, 327–341.

Received December 15, 1955.

PRIDE IN GROUP PERFORMANCE AND GROUP-TASK MOTIVATION[1]

LEONARD BERKOWITZ

University of Wisconsin

AND BERNARD I. LEVY

Crew Research Laboratory, Air Force Personnel and Training Research Center

INVESTIGATIONS conducted by the Survey Research Center (SRC) have shown that the attitude constellation variously labeled "pride in work group" (e.g., 4) or "pride in group performance" (7) tends to be related to group productivity. In discussing this relationship, investigators (e.g., 5) have suggested that work-group members develop pride in their group's ability to perform the designated task as a result of perceptions of past performance effectiveness, and that the resulting attitudes then contribute to future performance effectiveness.

The principal criterion employed in the SRC studies has been workgroup "productivity." It is plausible to assume that task motivation is among the major determinants of this variable. On this assumption, two aspects of the hypothesized relationship between pride in group performance and task motivation can be distinguished. In the first place, since attitudes toward the *group's* performance are explicitly at issue, the relevant favorable attitudes held by any individual member supposedly result from his perceptions of the group's effectiveness rather than of his own effectiveness alone. A measure of pride in group performance therefore should be affected more strongly by evaluations of the performance of the group as a whole than by evaluations of the separate group members, particularly when the individual members have no cues as to how well each of the others in their group is doing.

The second and major consideration in the hypothesized relationship is that pride in group performance is assumed to be associated with high motivation to perform the task assigned to one's group, and precedes later high productivity. To our knowledge there is little direct evidence on this last point. A study reported by the Crew Research Laboratory (6) has demonstrated that aircrews in training that have high "pride in crew" (which is conceptually similar to pride in group performance) tend to be rated highly by their superiors in later combat assignments. However, since the superiors' ratings may depend on factors other than task motivation (such as interpersonal harmony among the crew members) it is at least conceivable that this finding may not be generalizable to the relationship between group attitudes and task motivation.

The general purpose of the present experiment is to test the relationship between pride in group performance and task motivation under carefully controlled conditions. Subjects' (*S*s') "pride" in their group's performance was manipulated by means of arbitrary performance evaluations. The evaluations varied in their nature (whether they were favorable or unfavorable) and in their target (whether they were directed to the group as a whole or to each of the individual members). An analysis was conducted to determine the effects of the nature and target of the performance evaluations upon pride in group performance. The different conditions then were compared in terms of an index of task motivation. The hypothesis, derived from the SRC studies, was that groups receiving favorable evaluations of the group as a whole would score higher on this motivational index than groups receiving unfavorable evaluations of the entire group or groups in which the members receive evaluations of solely their own performance. The effect of subsequent task performance upon attitudes toward the group was finally investigated.

[1] This study was performed under the Air Force Personnel and Training Research Center, Lackland Air Force Base, San Antonio, Texas in support of Project 7713. Permission is granted for reproduction, translation, publication, use and disposal in whole and in part by or for the United States Government. The opinions expressed in this paper are those of the authors and do not necessarily reflect the views or have the endorsement of the U. S. Air Force or U. S. Government.

METHOD

Subjects. The present sample consists of 25 groups, each consisting of three airmen drawn from a population of basic trainees. The airmen varied markedly in intelligence classification but included substantial representation from the highest intelligence groups.

Group task. The group task was employed solely as a convenient setting within which the effects of performance evaluations could be studied. The assigned task was based upon Lanzetta's modification of a task first employed by Schutz,[2] involving a miniature 3-man Air Defense warning center. Three discrete target areas were located near the center of a metal-covered board with a grid identified by coordinates. Model planes with small attached magnets represented "enemy" and "interceptor" air forces. The following types of information were supplied to each group: (*a*) A continuous tape-recorded message, divided into approximate 4-minute intervals, reported the coordinates of all planes (other than friendly interceptors): 15 unidentified aircraft, 9 enemy bombers and 6 friendly planes. During each interval the group had the task of translating the aircraft positions reported into moves on the interceptor board, and deciding on the deployment of interceptors. (*b*) A second input consisted of flight plans for the 6 friendly planes posted on the "Friendly Aircraft Status Board" near the metal-covered board. The general requirement was to defend the three target areas against an attack by enemy bombers. To do so necessitated correct identification of friendly and enemy planes, and strategic deployment of the available interceptor force for counterattack.

Procedure. All *S*s first were told that the purpose of the experiment was to study the way in which people work together in important Air Force groups such as Air Defense teams. Detailed instructions were given about the sequence of activities which included a brief practice session, an initial trial, a 15-minute "break" period, and finally a second trial. The *S*s in the experimental groups were advised that psychologists and experts would observe their performance from behind a one-way mirror and evaluate them by comparing their performance with that of over 200 groups of basic trainees like themselves who had worked previously on the task. They were informed that they would receive psychological evaluations at the end of the fourth and eighth intervals of the first trial. Nothing was said to the control groups about their being evaluated. An "objective" scoring scheme was then explained to both experimental and control *S*s: they would receive 20 points for every enemy bomber downed, lose 50 points for every target hit, and lose 20 points for every friendly plane downed and fighter lost.

After a brief practice trial of two intervals, the *S*s were given the first trial, which was stopped before the end of the problem, at the completion of the eighth interval. The *S*s then filled out a brief attitude and sociometric questionnaire, after which they were told they could take a 15-minute "break" but that they would have to remain in the room.

The *E* then went to the adjoining room where he observed the *S*s' behavior during this rest period through a one-way mirror. Magazines were displayed on a table at one end of the room for the *S*s to read if they desired, although nothing was said to them about this. The *E* categorized their behavior and discussions.

At the end of 15 minutes the *E* returned to the experimental room and started the *S*s upon the second trial. Upon completion of the problem the *S*s were told how many enemy bombers had been downed and how many targets had been hit. The questionnaire then was readministered and the *S*s dismissed.

Experimental manipulations. There are five conditions of five groups each. Four of the conditions represent a factorial design with the following sources of variance: (*a*) nature of evaluations; (*b*) target of evaluations; (*c*) interaction of nature and target; and (*d*) residual. Half of the experimental groups received favorable evaluations of their performance (+) during the first trial while the other half of the sample received unfavorable evaluations (−) during this trial. Similarly, half of the experimental groups received evaluations of the group as a whole (G); the individual members were not singled out and each member saw the group rating. The *S*s in the remaining evaluated groups (I) received ostensible evaluations of their own individual performance. The *S*s in the I conditions were instructed not to communicate with each other concerning their ratings. Evaluation of both groups and individuals was in terms of percentile scores on ratings of seven attributes such as: (*a*) understanding of problem requirements; (*b*) utilization of resources; (*c*) coordination; and (*d*) timing and pacing of work. It was believed that these attributes would seem meaningful to the *S*s but yet would be nebulous enough so that the *S*s would not have any objective basis for questioning the ratings. The ratings given the praised groups averaged 85 at the fourth interval report and 90 at the eighth interval. For the criticized groups the ratings averaged 35 and 30 respectively.

The fifth condition was a control (C), included in order to examine the effects due to the evaluations per se, regardless of their nature. Groups in this condition were not told that they were to be observed and evaluated and were not given any ratings.

Observations. The observations made by the *E* during the 15-minute "break" period consisted of 15-second time samples taken each minute. Recording was in terms of five categories: "Keeps to self" (S); "Task-oriented discussion" (T-O); "Non-task-oriented discussion" (N-T-O); "Critical remarks" (C); and "Remarks concerning the present experiment" (i.e., what the purpose was, what the experimenter would do next, etc.) (E). No critical remarks were ever heard. Since very few remarks were coded in Category E, this category was not used in the analysis.

The observers coded any *S* reading a magazine or otherwise not engaged in conversation during the time sample as "keeping to himself." A remark made at any time during the 15-second period was coded in one of the three verbal categories according to its predominant function. For the present analyses, the total number of units coded in any one category over all

[2] We are indebted to Dr. John Lanzetta for the construction of the Air Defense task and for suggestions for the modification of the task for the present experiment.

15 time samples are employed as group scores. The number of units coded as task-oriented discussion (T-O) is taken as an index of the group's task motivation. This measure, the major dependent variable, is an index of the group members' concern with their assigned task at a time when they are under little overt pressure to be so concerned.

An estimate of the reliability of observations was computed using data from five randomly selected groups observed by both *E*s, involving an analysis of variance design in which the classifications were categories, observers, and groups. The continuous variable was the group score in each category. The Hoyt (1) reliability estimate from this analysis was .99.

Attitude and sociometric measures. The attitude questionnaire administered immediately after each trial contained one item designed to test the success of the experimental manipulations, and nine items to assess pride in group performance. Five of the nine items were selected for the pride-in-group scale after applying the iterative scaling technique described by Stice and Knoell (6) to the post-first-trial responses. All analyses were based upon group scores on these five items. In combining the items, scores on each item first were converted into *T* scores. The group-scale score was the sum for all members over all items. A constant of 200 was subtracted from each group-scale score to facilitate the analysis. Two of the items in the scale are:

> If you were going to go through this exercise again, would you expect to do better than: (Check one)
>
> ____20 per cent of the other people going through it
> ____40 per cent of the other people going through it
> ____60 per cent of the other people going through it
> ____80 per cent of the other people going through it
> ____All others.
>
> How well do you think each of your present teammates has worked on the problem you have just completed? (Poor job, 1, through Excellent job, 5.)

Objective performance measures. The objective measures employed here were number of "enemy bombers" downed before hitting a target, and the number of "friendly aircraft" accidentally shot down. We believe that these performance measures are dependent largely upon intellectual and problem-solving processes rather than group-task motivation. The performance data were collected, not as criteria or for validation of the motivation index, but in order to study the effects of quality of performance during the second trial upon subsequent changes in pride in group performance.

RESULTS

Success of the experimental manipulations. The questionnaire item designed to test the success of the manipulation reads as follows:

> What is your impression of your group's effectiveness as an Air Defense team as compared with the effectiveness of other 3-man teams composed of basic trainees going through this exercise? (Check one):
>
> ____Better than 20 per cent of the other teams
> ____Better than 40 per cent of the other teams
> ____Better than 60 per cent of the other teams
> ____Better than 80 per cent of the other teams
> ____Better than all others.

The weight given to each alternative was the percentage of other teams thought by each *S* to have been less effective than his team. The score for each group was obtained by summing the scores for each of the three members. The mean group score immediately after the first trial in each of the five conditions is shown in Table 1.

Analysis of variance of the data in the four experimental conditions suggests the following: there is no significant difference between the perceived effectiveness scores of the G groups as compared with the I groups (F = 0.2). However, the favorably evaluated groups, as would be expected, have much higher scores on this measure than the unfavorably evaluated groups (F = 1,071.6). There is also a significant interaction (F = 6.69; p = .05), which may be a function of the amount of information provided to the *S*s concerning the performance effectiveness of others in the group.

In comparison with the control groups' scores, the groups in the two favorably

TABLE 1

MEAN PERCEIVED GROUP-EFFECTIVENESS SCORES IN EACH OF THE FIVE CONDITIONS

Target	Nature of Evaluations: Favorable (+)	Unfavorable (−)
Entire group (G)	220	100
Individual members (I)	192	120
Control condition:	144	

TABLE 2

MEAN "PRIDE-IN-GROUP" ATTITUDE SCORE AFTER THE FIRST TRIAL IN EACH OF THE FIVE CONDITIONS

Target	Nature of Evaluations: Favorable (+)	Unfavorable (−)
Entire group (G)	58.93	45.82
Individual members (I)	54.69	43.60
Control condition:	45.39	

evaluated conditions had perceived effectiveness scores significantly greater than the control condition ($t = 5.07$, $p = .001$, for the G+ groups; and $t = 3.20$, $p = .01$ for the I+ groups), while the G− groups had perceived effectiveness scores significantly lower than the control groups ($t = 2.93$, $p = .05$).

Pride in group after the first trial. Table 2 presents the mean pride-in-group score in each of the five conditions for the responses made immediately after the first trial. Analysis of variance of the data for the experimental conditions shows that the favorably evaluated groups tend to have significantly higher attitude scores than the unfavorably evaluated groups ($F = 45.96$, $p = .001$). The main effect for target, G or I, while not significant ($F = 3.29$, $p = .10$), indicates a trend in the direction of our hypothesis.

The t test of the difference between the G+ condition and the pooled groups in the three other conditions (using a one-tailed test)[3] is highly significant ($t = 5.34$; $p = .001$). This condition also has a significantly higher score than the next highest condition, I+ ($t = 2.08$; $p = .03$). Thus, the pride in group performance measure, as predicted, is more strongly affected by the perceived merit of the total groups's performance than by each member's perception of his own individual performance. Lastly, groups in both favorably evaluated conditions tend to have somewhat higher pride-in-group scores than the control groups, while there are no significant differences between the control and unfavorably evaluated groups.

Behavior during the "break" period. The observers' frequency tallies were found to have a skewed distribution and, therefore, were normalized by the square-root transformation. The mean transformed scores for each of three categories are given in Table 3. It can be seen from this table that the ordering of conditions in terms of pride in group performance is not the same as their ordering on this measure of task motivation, although the G+ groups are high on both variables.

Analyses of variance of the data in the four experimental conditions indicate that the target of the evaluations has a significant effect upon the T-O and S categories of behavior ($F = 4.52$, $p = .05$ and $F = 4.89$, $p = .05$ respectively) but not on the N-T-O category. Task-oriented discussion is more frequent when the evaluations—favorable or unfavorable—are given to the group as a whole, and more *S*s keep to themselves when evaluations are given to individual group members. The level of motivation on the present interdependent group task apparently is influenced by whether evaluation is of the group as a whole or of each individual's own performance. Furthermore, the high frequency of *S*s keeping to themselves in the I conditions suggests that individual evaluation tends to divide the group.

We had hypothesized that the groups in the G+ condition would have the highest scores of all conditions on the task-motivation index. This hypothesis was tested by t tests of the difference between this condition's means on the observational categories and the means obtained by pooling the groups in the other three conditions, basing the error terms upon the residual mean squares. The resulting t's are significant for categories T-O and S. They are 1.92 ($p = .04$) and 1.75 ($p = .05$) respectively. Thus, the G+ groups tend to have more task-oriented discussions and fewer instances of members keeping to themselves than the 15 groups in the three other conditions.

The t tests were also made to determine if the evaluated experimental conditions differed from the nonevaluated control conditions in the frequencies of behavior in the three categories. There were no significant differ-

TABLE 3

TOTAL NUMBER OF BEHAVIORAL UNITS IN EACH OF THE THREE CATEGORIES DURING THE "BREAK" PERIOD

(Square-root transformation)

Target	Nature of Evaluations					
	Favorable (+)			Unfavorable (−)		
	T-O*	N-T-O†	S‡	T-O	N-T-O	S
Entire group (G)	23.4	13.7	12.6	18.6	16.7	15.2
Individual members (I)	13.9	18.3	19.2	10.9	14.9	23.1
	T-O		N-T-O		S	
Control condition:	14.2		12.1		24.6	

* Task-oriented discussion
† Non-task-oriented discussion
‡ Keeping to himself

[3] All t tests specified by the a priori hypotheses employ one side of the distribution.

ences between any of the evaluated conditions and the control condition for the T-O category. A significant tendency is found, however, for the G+ condition to result in fewer instances of members keeping to themselves (S) than the control condition, a finding consistent with the interpretation that favorable evaluation of the group as a whole has served to unite the members of the group.

In summary, the tendency for the G+ group to be different from the groups in the other conditions on both pride in group performance and degree of concern with the task suggests that there may be two sets of factors involved in the present attitudes related to motivation to perform the assigned group task. The high pride that is associated with high group-task motivation apparently results from perceptions of (*a*) effective performance, (*b*) by the group as a whole.

Objective performance measures for the second trial. There were no statistically significant differences in group performance during the second trial. As explained earlier, performance on this Air Defense task was thought to be relatively independent of task motivation.

Pride in group after the second trial. The ordering of conditions in terms of the post-second-trial attitude scores is not similar to the post-first-trial ordering, and there are no significant differences among the five conditions. There are also no significant effects for the four experimental treatments or their interaction.

A comparison of the scores obtained in the different conditions before and after the second trial indicates that the unfavorably evaluated groups had increased in their "pride" scores, while the favorably evaluated groups had generally shown a decrease in these scores. An attempt was made to determine if the change in the attitude scores was related to the initial attitude level, later performance, or the interaction of these two variables.

The distribution of scores on each of the two variables, post-first-trial pride in group and second-trial performance effectiveness, was first dichotomized. (The performance measure was obtained employing the weighting scheme described to the *S*s). The 25 groups then were divided among the four conditions of the resulting factorial design, omitting three groups selected at random from the Low First-Trial Attitude–High Performance condition in order to obtain proportional cell frequencies. The mean change in attitudes from first to second trial in each of the four conditions is given in Table 4.

TABLE 4

MEAN CHANGE IN PRIDE IN GROUP IN EACH OF THE FOUR COMBINATIONS OF POST-FIRST-TRIAL PRIDE IN GROUP AND SECOND TRIAL PERFORMANCE EFFECTIVENESS

Performance Effectiveness	Post-First-Trial "Pride in Group"	
	High	Low
High		
N	6	5
Mean change	−4.1	5.3
Low		
N	6	5
Mean change	−5.6	4.6

The data, arranged in the design shown in Table 4, were subjected to an analysis of variance. Two of the three effects are statistically significant. 1. The groups with initially high attitudes tend to develop less favorable attitudes while the groups with lower attitude scores tend to develop more favorable attitudes ($F = 9.91$, $p = .01$). This regression effect suggests that the pride in group performance brought about by the present experimental manipulations is relatively unstable. 2. The interaction of initial attitude score and performance also is significant ($F = 9.77$, $p = .01$). In other words, the groups with initially high pride in group but relatively poor later performance tend to decrease the most in their attitude scores, and the groups with initially low pride in group but relatively good later performance show the greatest increase in attitude scores.

DISCUSSION

It has been assumed that the amount of task-oriented discussion within a group during the work groups' "break" period provides an index of group-task motivation. The groups were under little overt pressure to be concerned with their task during this rest period and the presence of magazines in the room implied that they were free to read or engage in other non-task-related activities. Furthermore, since there were no between-condition

differences in objective performance prior to the "break" period, differences in the amount of task-oriented discussions can safely be attributed to the performance evaluations given to each condition. The obtained results, then, support the implications inherent in the SRC hypothesis. The pride in group performance associated with high group-task motivation results from group members' perceptions of *group* effectiveness rather than from each member's perception of his own effectiveness alone.

What processes produce the relatively high group-task motivation under conditions of favorable group evaluation? Katz and Kahn (5) suggest that the high group-task motivation associated with pride in group performance is a direct result of increased group attractiveness. However, experimental evidence indicates that a high level of group attractiveness can be as readily associated with low group-task motivation as with high motivation depending upon such other factors as group standards for low or high productivity (1, 8). In those groups having a standard for high production, groups highly attractive to their members should be more highly task-motivated than less attractive groups. The attainment of the group goal of high productivity by the former groups, furthermore, may well result in relatively high pride in group performance. But a problem remains. Why do some groups develop standards for high production and not others?

A better explanation, we believe, can be found by considering the two aspects of the evaluations, their nature and their target. Favorable evaluations should reward the group members, but differences in the target of the evaluations result in different things being rewarded. If the *individual* members are favorably evaluated, they receive rewards for their *individual* performances. If the group is favorably evaluated each *S* receives a reward only by virtue of his being a member of the group. In other words, the favorable evaluations of the entire group serve to create the perception that the members are interdependent in attaining rewards; each *S*'s reward attainment is dependent upon the others in his group. Research findings indicate that this type of perception can produce relatively high group-task motivation. Deutsch (2) experimentally established "cooperative" and "competitive" group situations, differing in the extent to which experimental conditions favored the mutual perception of "promotive interdependence" among the members. Members of the "cooperative" groups felt stronger obligation to each other and were more highly motivated to work hard at the group task than did the "competitive" group members. The relationship between pride in group performance and task motivation may result, thus, from a process in which group members high in pride perceive that each member's attainment of the rewards associated with effective performance is dependent upon the others in the group.

Two sets of evidence obtained by the present investigators support the contention that the target of the performance evaluations can affect the members' perceptions of interdependence. First, the "individual evaluations" condition in the present experiment has a greater incidence of members keeping to themselves in the "break" period than the condition receiving "group evaluations," while the G+ groups had the smallest amount of this type of behavior. Although alternative explanations can be found, it is plausible to assume that group members perceiving a high degree of member interdependence for goal attainment would be relatively unlikely to keep apart from the others in the group. Second, a yet unpublished study compares the condition in which members are aware that everyone in the group has received favorable individual evaluations with the condition in which favorable evaluations are given to the group as a unit. As was argued above, members of groups in the "individual evaluation" condition should not perceive the attainment of their rewards as being greatly dependent upon the others in the group even though they may be aware that everyone has done well. The results are in accord with this hypothesis. The former groups had a smaller incidence of task-oriented discussions and a greater incidence of members keeping to themselves than the groups favorably evaluated as a unit.

SUMMARY

Previous studies have found that work groups having high pride in group performance

tend to be relatively highly productive. It generally is assumed that the attitudes have contributed to the effective performance. The present experiment was designed to test implications of this hypothesized relationship, namely, that pride in group results from group members' perceptions of (*a*) effective task performance of (*b*) the group as a whole (rather than of their own separate performances) and is directly associated with high motivation toward the assigned group task.

Groups in the present study were assigned to work on a simulated Air Defense task. They received an initial trial, on the basis of which they supposedly were evaluated, a "break" period, and then a second trial. Twenty-five 3-man groups were distributed among five conditions: two conditions in which group performance on the initial trial was favorably evaluated, and two conditions in which this performance was unfavorably evaluated. One of the favorably evaluated and one of the unfavorably evaluated conditions received evaluations of the group as a whole while the other conditions received evaluations of the individual members. In each case the evaluations were provided through previously constructed ratings. A final, nonevaluated condition served as a control.

The amount of task-oriented discussion during the "break" period was used as the measure of task motivation. Groups receiving favorable evaluations of the group as a whole were found, as predicted, to have higher pride in group performance and greater concern with the assigned group task than groups having either unfavorable group evaluations or evaluations, whether favorable or unfavorable, of the separate group members.

Evidence is cited supporting the hypothesis that the high group-task motivation related to high pride in group performance results from a perception of interdependence among the group members with respect to the attainment of reward. Conditions in which group members presumably believe they can be rewarded for effective performance somewhat independently of the others in the group result in a relatively high incidence of members keeping to themselves during the "break" period and in relatively low group-task motivation.

The attitude responses made after performance on the second trial yielded the following results: The level of a group's pride in group established at the termination of the first trial was relatively unstable. There was a significant tendency for the groups with highly favorable attitudes to develop less favorable attitudes, and for the groups with low scores to increase in their "pride." The level of pride in group changed if the initial level was at variance with the quality of the group's performance, an effect that may be dependent, in part at least, upon the relative instability of the initial attitude level.

REFERENCES

1. BERKOWITZ, L. Group standards, cohesiveness and productivity. *Hum. Relat.*, 1954, **7**, 509–519.
2. DEUTSCH, M. The effects of cooperation and competition upon group process. *Hum. Relat.*, 1949, **2**, 199–231.
3. HOYT, C. J. Test reliability estimated by analysis of variance. *Psychometrika*, 1941, **6**, 153–160.
4. KATZ, D., MACCOBY, N., & MORSE, NANCY. *Productivity, supervision and morale in an office situation*. Ann Arbor: Survey Research Center, Univer. of Michigan, 1950.
5. KATZ, D., & KAHN, R. L. Some recent findings in human-relations research in industry. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1952. Pp. 650–665.
6. KNOELL, DOROTHY M. Relationships between attitudes of bomber crews in training and their attitudes and performance in combat. *USAF, Personnel Train. Res. Cent., Res. Bull.*, in press.
7. MORSE, NANCY. *Satisfactions in the white collar job*. Ann Arbor: Survey Research Center, Univer. of Michigan, 1953.
8. SCHACHTER, S., ELLERTSON, N., MCBRIDE, DOROTHY & GREGORY, DORIS. An experimental study of cohesiveness and productivity. *Hum. Relat.* 1951, **4**, 229–238.
9. STICE, G. F., & KNOELL, DOROTHY M. A simple mean-difference technique for obtaining scales. *USAF Hum. Resour. Res. Cent., Res. Bull.*, September 1953, No. 53–36.

Received December 19, 1955.

EFFECTS OF WORK-GROUP STRUCTURE AND CERTAIN TASK VARIABLES ON GROUP PERFORMANCE[1]

JOHN T. LANZETTA

Fels Group Dynamics Center, University of Delaware

AND THORNTON B. ROBY

Institute for Applied Experimental Psychology, Tufts University

ONE feature characterizing many complex man-machine systems is the fact that relevant information regarding the condition of the system, which is necessary for various control actions, may not be available at a single vantage point. Even if readily accessible, the volume of information to be processed is generally so large as to make it virtually impossible for a single individual to monitor more than a small fraction of the relevant input displays. Since decisions and control actions typically require information about the status of many system variables, such systems are of necessity manned by teams in which each individual serves both as an operator, or control agent, and as an observer, or information source for himself and other control agents. Where face-to-face interaction is impossible this requires a communication network so that each operator's information can be supplemented by information relayed by other group members.

It is a plausible hypothesis that in systems of this sort performance will be affected by the volume of information to be relayed, the extent to which relevant information is dispersed throughout the group, and the nature of the communication "net" supplied for information transmission. The latter variable has received considerable research attention in recent years (e.g., 9) but the effects on group performance of variation in volume and dispersion of information have received little systematic exploration. The present study is one in a series directed toward investigation of these particular variables.[2]

A previous study (7) examined the performance of three-man groups operating under four imposed communication structures. Each of the structures was symmetrical in that all three group members had identical task requirements, and all entailed the processing of the same amount of information and the execution of equivalent control adjustments. The structures differed only in the proportion of information which had to be transmitted and in the extent to which relevant information was dispersed over group members. The principal finding of this study was that performance could be related in a meaningful way to the volume of information which had to be transmitted and to the nature of the sources of the information on which control decisions were based. If relevant information was directly available to a response "agent" in the form of a visual display, or if relayed information could be obtained from a single outside source, performance on controls was markedly better than when information had to be obtained from several external sources.

The present study is intended to extend this investigation of structure while at the same time exploring the importance of certain task variables. Considerable research has been directed to the determination of the information-handling capacity of individuals (e.g., 5), but little parallel research has been done on teams. How does team performance vary as a function of information processing demands? Does the information-handling capacity of teams depend on the requirements for internal transmission of information imposed by a communication structure? Both of these questions are of interest in the present investigation.

[1]This research was performed in support of ARDC Project 7713 under the Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government.

[2]A discussion of the framework on which the studies are based is presented in an unpublished paper; J. T. Lanzetta, and T. B. Roby, "A framework for the study of work-group structure and task parameters," Laboratory Note, Crew Research Laboratory, LN-55-9, June 1955, and a brief discussion of the general theoretical model underlying the approach is presented in a forthcoming publication (8).

The two structures employed in the study were selected from the four used in the previous experiment (7). They represent the conditions which differed most markedly in the extent to which they required information transmission. The two task variables involved operations on the input to the group. A differential and varying characteristic of many group tasks is the amount and predictability of the inputs which must be handled. Both the rate of change of the input and the predictability of the locus of a change were varied in the present study.

METHOD

Apparatus and Task

Since the apparatus and general features of the task have been previously described in detail (7), they will be summarized briefly here. Each of the three *S*s belonging to a group was seated in a separate booth which prevented visual and direct auditory communication between group members. Each booth contained a standard aircraft interphone circuit with a throat mike and headset, and all communication between group members was by means of this circuit. By depressing a hand switch, any *S* could communicate simultaneously with both other *S*s.

In each booth there were two switches, each having an "off" setting and three "on" positions. Near each switch was a printed table of operating instructions stating the correct switch settings for various combinations of instrument readings. A LaBelle automatic slide projector behind each *S* was used to project a picture of two simulated aircraft instruments on the front wall of the booth. The projectors permitted stacking of about 30 slides which were tripped into the projection chamber by a remote-control switch operated by *E*. The pairs of simulated instruments presented to the *S*s were different in each booth.

The correct setting for each control could be determined only by reference to joint readings on two instruments. Conversely, each instrument reading was relevant to the correct settings for two controls (see Table 1). The general task set the group was to process the information presented (instrument readings), relaying the necessary information to the proper booths, and to execute control actions based on relayed or directly available instrument readings. Each of the three "on" positions for a given switch was connected to one pen on a 20 pen Esterline-Angus recorder so a complete time record of the six control settings made by the group could be obtained.

Experimental Conditions

Structure. Two structure conditions were employed. In the Low Autonomy Structure (LA) each *S* had two controls for which none of the four necessary instrument readings was directly available. In the High Autonomy Structure (HA) each *S* had two controls for which all but one of the requisite instrument readings were available in his booth. Thus, LA required each response agent to obtain four units of information from external sources, while HA required each *S* to obtain one unit from external sources. In addition, in the LA Structure the information required by each member was distributed among both other group members while in the HA Structure the required information was available from a single external source. The interdependencies between group members are represented in Table 2 by a matrix schema.

With the apparatus employed it was most convenient to fix the location of instruments within a booth for all conditions. Thus, slides showing Compass Heading and Air Temperature were presented in Booth 1; Generator Voltage and Fuel Pressure in Booth 2; and Air Speed and Rate of Climb in Booth 3. It also seemed advisable to maintain the same fixed dependencies between control settings and instrument readings for all conditions as indicated in Table 1. Hence structure was manipulated by changing the location of the controls and involved only the moving of control name plates and instruction cards from one booth to another. It can be readily seen from examination of Table 1 that with a fixed allocation of instruments changing the location of controls varies the required degree of communication.

Task variables. Three input parameters were ma-

TABLE 1

INFORMATION BASES OF EACH OF THE SIX CONTROLS

Control	Relevant Information					
	O_1 Compass heading	O_2 Air temperature	O_3 Generator voltage	O_4 Fuel pressure	O_5 Air speed	O_6 Rate of climb
C_1. Power setting	x	x				
C_2. Steering mechanism	x			x		
C_3. Landing gear			x	x		
C_4. Control switch			x			x
C_5. Selector knob					x	x
C_6. Reset lever		x			x	

Note.—An "x" in a given row and column indicates that a reading on the instrument at the column heading is relevant to decision for the control in the row.

TABLE 2

INFORMATION TRANSMISSION REQUIREMENTS IN THE TWO EXPERIMENTAL STRUCTURES*

	Source of Information			
	Group Member	A	B	C
High Autonomy Structure (HA)				
User of information	A	3	1	0
	B	0	3	1
	C	1	0	3
Low Autonomy Structure (LA)				
User of information	A	0	1	3
	B	3	0	1
	C	1	3	0

* The entries indicate the items of information (instrument readings) held by the group member indicated in the column and required by the individual indicated in the row. Entries in the diagonal indicate items directly available. For correct control adjustments each of these items of information had to be transmitted for every sequence of three slide changes.

nipulated. The input to the group consisted of slide changes which altered the instrument readings within a booth. Nine different slides (two different instrument readings per slide) were presented in each booth by means of the LaBelle projector. The order in which slide changes were made in the three booths was staggered so that the total information display for the group changed 27 times, e.g., a slide was changed in Booth 1 then, a fixed interval later, in Booth 3, etc. The interval between successive slide changes, irrespective of booths, was one of the variables manipulated. Changes were made at either 15-second or 10-second intervals.

The second independent task variable was the predictability of the change sequence. The order in which slide changes were presented in the booths was either "random" (subject to the restriction that each booth received a slide change before the next sequence was initiated), or "predictable." In the former case the order of slide changes for the three booths was determined from a table of random numbers, in the latter case the order of change was always Booth 1, Booth 2, Booth 3.

The third task variable was of peripheral interest and was included primarily as a methodological check. The 30 slides for each booth were arranged in two different sequences resulting in different required patterns of control changes.

Subjects and design. The *S*s were Air Force basic trainees undergoing training at Lackland Air Force Base. Six *S*s were available for each experimental day; the six were divided fortuitously into two groups of three *S*s each. The experimental conditions were randomly ordered prior to arrival of *S*s, and groups were assigned to a condition in the order in which they were received. Pertinent demographic characteristics of the *S*s were not available. However, the age range of this population is generally from 18-22 years and the range of education and intelligence is larger than that in most college samples.

The three task variables were arranged in a 2 × 2 × 2 factorial design with two groups assigned to each cell. Each of the two groups performed under both structure conditions, the order of presentation of the structure conditions being counterbalanced for each cell.

Administration

The *S*s were brought into the experimental room, seated in the booths fortuitously, and given a brief introductory talk. They were informed that the experiment would not affect their Air Force status and that the experimental task required them to operate much like members of a bomber crew. It was explained that—as in a bomber—they would make certain control adjustments on the basis of the best available information but that *all* the information usually would not be available to any one of them, so they would need to obtain information from other crew members. After they had been shown how to adjust and use the interphone equipment, they were given three practice slides. No standard practice period was adopted, but the experimental sequences were not begun until the expermenter felt that all group members understood the task.

The information input series described was repeated three times for each group under both conditions. Between repetitions of the series (within conditions) a blank slide appeared, and all *S*s reset their controls to "off." At the conclusion of the three periods of the first session, the men were permitted to take a 10- to 15-minute break while the control labels were changed. Thus the total running time for each group was about 20 minutes for the introductory talk; 21 minutes for each of two experimental sessions; and 10 minutes for changing apparatus.

RESULTS AND DISCUSSION

The scores used in the present analysis are the number of errors within three periods, each period consisting of 27 slide changes. Each time a slide was changed in any of the three booths a new set of control adjustments was required: if a correct adjustment was not

TABLE 3

GROUP ERROR SCORES FOR EXPERIMENTAL CONDITIONS

Predictability	Slide Sequence	Structure Sequence	Rate of Change			
			10-second		15-second	
			1	2	1	2
Random	S.S. I.	AD	347	337	168	33
		DA	174	219	62	162
	S.S. II.	AD	278	184	229	64
		DA	242	247	257	247
Predictable	S.S. I.	AD	296	115	130	45
		DA	170	217	42	66
	S.S. II.	AD	208	203	69	199
		DA	225	260	83	164

TABLE 4
SUMMARY OF ANALYSIS OF VARIANCE: GROUP ERROR SCORES

Source of Variance	*df*	Mean Square	*F* Ratio
1. Total	31	13,239.660	
2. Between groups	15		
A. Rate of change	1	5,029.174	20.479 ($p < .01$)
B. Predictability	1	997.507	4.062
C. Slide sequence	1	576.000	2.346
D. Structure sequence	1	8.028	
A × B	1	14.063	
A × C	1	693.444	2.824
A × D	1	225.000	
B × C	1	12.250	
B × D	1	.111	
C × D	1	733.507	2.987
Pooled higher order interaction	3	245.571	
3. Within groups	16		
E. Session	1	82.507	2.432
F. Structure	1	1,320.111	53.757 ($p < .05$)
E × A	1	16.764	
E × B	1	166.840	4.917
E × C	1	51.361	
F × A	1	1.361	
F × B	1	81.000	
F × C	1	253.340	7.467
E × A × B	1	291.840	8.601
E × A × C	1	14.694	
E × B × C	1	850.694	25.072 ($p < .05$)
F × A × B	1	342.250	10.087
F × A × C	1	15.340	
F × B × C	1	166.840	4.917
Pooled higher order interactions	2	33.934	

Note.—Homogeneity of the variances attributable to higher order between and within interactions, respectively, was tested by Bartlett's test (2). The hypothesis could not be rejected in either case.

made before the next slide change occurred, an error was recorded for that control. The error score for a group, then, consisted of the number of errors for the six controls for the 81 slide changes.

Table 3 presents the error scores for the experimental conditions. The results of analysis of variance of these data are presented in Table 4.

Differences for two of the three experimental variables of primary interest are significant at the .01 level. Error scores are lower for the longer (15-second) interslide interval and for the structure condition requiring the least amount of information transmission (HA). The "predictable" order of slide changes did not result in significantly lower error scores than the "random" order although the differences were in the expected direction. The interactions between task variables or between structure and task variables were not significant.

Although not presented here, a further analysis of scores for the individual control adjustments and periods indicated consistency of group performance during periods as well as over-all improvement in performance over periods. Similar results had been obtained in the previous study (7).

Task Conditions

Examination of errors for the four task conditions indicates an ordering of scores over conditions. The 10-second random condition appears to be the most "difficult," with the 10-second predictable, 15-second random, and 15-second predictable representing decreasing levels of "difficulty." It is apparent that the 10-second rate of change necessitates the processing of more information per unit time than the 15-second rate and that the predictable condition presents cues as to the sequence of information changes not presented under the random condition. Thus, both the rate of change and predictability variables modify the number of equally likely information states per unit time among which the group must discriminate.

A method of measuring the "information" or "uncertainty" of a source is available from information theory (5, 6). Applied to the present variables, the average amount of information presented a group in bits per minute for the four conditions is as follows:[3] 10-second

[3]There were two instruments per slide and three possible readings on each instrument, thus 3^2 equally probable information alternatives for each slide. The information (in bits), per slide is, therefore, equal to the $\log_2 3^2$. At the 15-second rate of change the information in bits per minute would be four times this value, while for the 10-second rate it would be six times this value. When slide changes occur in a random order however, the equiprobable states at each slide change consist of the number of alternatives for each of the slides (3^2) multiplied by three, since each slide has an equal chance of being presented. The information in bits per slide change is thus equal to the $\log_2 3 \times 3^2$. For the predictable sequence all 27 possible alternatives are available for the first change, 2×3^2 for the second change, and 3^2 for the third and all subsequent changes since the order of slide changes is established after three changes. The information in bits per minute for the 10- and 15-second rates of change for the random order than are 6 $\log_2 27$ and 4 $\log_2 27$ respectively. The information in bits per minute for the 10- and 15-second rates for the predictable order are approximately 6 $\log_2 9$ and 4 $\log_2 9$ respectively. In the latter

TABLE 5

MEAN GROUP ERROR SCORES FOR STRUCTURE AND TASK CONDITIONS

	Rate of Change			
	10-second		15-second	
	Random	Predictable	Random	Predictable
Low autonomy	272.75	245.25	201.50	107.25
High autonomy	234.25	178.25	104.00	92.25

random, 24.12; 10-second predictable, 18.96; 15-second random, 16.08; 15-second predictable, 12.64. It is evident that error scores are an increasing monotonic function of the amount of information per unit time presented the group.

Previous results obtained with individuals have indicated a relationship of essentially similar form between information rate and performance criteria (3, 4). However, even at the minimum input rate for our groups (12 bits per minute) a considerable number of errors were made. For individuals performing simple discrimination tasks a capacity of approximately 5 to 20 times this rate has been established (5, 6). The discrepancy presumably reflects inefficiency in the systems arrived at by the groups for processing the presented information, and probably would not be so large if "optimum" procedures for routing information were used. However, further study on this point is needed.

Structure Conditions and "Linkage Types"

Table 5 presents the error scores for the LA and HA structures for the four task conditions. Error scores for both structures increase monotonically from the 15-second predictable treatment to the 10-second random treatment and thus, according to the above analysis, increase as a function of the amount of information presented per unit time. For both structures errors are essentially a negatively accelerating function of amount of information. However, there is an indication that differences between the two structures are maximal for the two intermediate values of input uncertainty (Fig. 1).

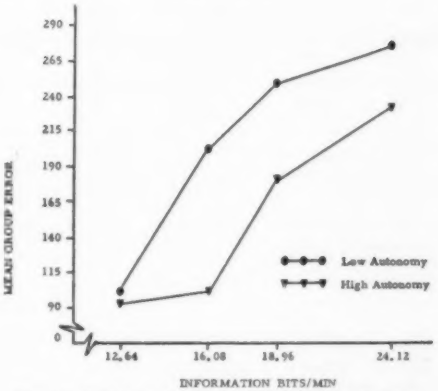

FIG. 1. RELATIONSHIPS BETWEEN MEAN GROUP ERROR SCORES AND INFORMATION RATE FOR THE TWO EXPERIMENTAL STRUCTURES

The lesser requirements for coordination in the HA structure would lead one to expect that the relative advantage of HA over LA would increase as the amount of information presented to a group increased. Examination of the results for specific controls may shed some light on the unexpected relationships. The previous study (7) had indicated that much of the variance attributable to structure could be accounted for in terms of "linkage types."

Classification of controls in terms of whether information was directly available or relayed and whether it was located at a single source or dispersed resulted in four qualitatively difference linkage types. The four types are: controls for which the two items of informa-

TABLE 6

MEAN ERROR SCORES FOR FOUR TASK CONDITIONS BY LINKAGE TYPE

	Linkage	Rate of Change			
		10-second		15-second	
		Random	Predictable	Random	Predictable
High autonomy	Sole source—direct	119.33	98.66	43.66	35.00
	Dispersed source—partially relayed	193.00	139.00	95.00	88.00
Low autonomy	Sole source—relayed	145.00	142.00	113.33	33.33
	Dispersed source—relayed	218.66	189.66	155.33	109.66

case, the information values were actually computed by averaging the information per slide change over the 81 slide changes since, as indicated above, the number of equiprobable alternatives is not equal to 3^3 for the first two changes.

TABLE 7

SUMMARY OF ANALYSIS OF VARIANCE: LINKAGE-TYPE ERROR SCORES

Source of Variance	*df*	Mean Square	*F* Ratio
1. Linkage type	3	18,590	22.518 ($p < .001$)
A. Sole vs. dispersed	1	39,331	47.642 ($p < .001$)
B. Structure	1	16,354	19.809 ($p < .01$)
A × B	1	86	
2. Source uncertainty	3	24,268	29.396 ($p < .001$)
3. Linkage type × uncertainty	9	832	
4. Within (error)	32	826	

tion required were directly available (sole source—direct); controls for which the two items of information required were available from *one* of the other group members (sole source—relayed); controls for which one item of information was directly available and one item was available from another group member (dispersed source—partially relayed); controls for which one item was available from each of the other group members (dispersed source—totally relayed).

Table 6 presents the error scores for the four task treatments by linkage type. The results of analysis of variance, shown in Table 7, indicate that variance attributable to linkage types is significant as is that attributable to whether the information is dispersed or available at a single source. Structure and "uncertainty of the input" are also highly

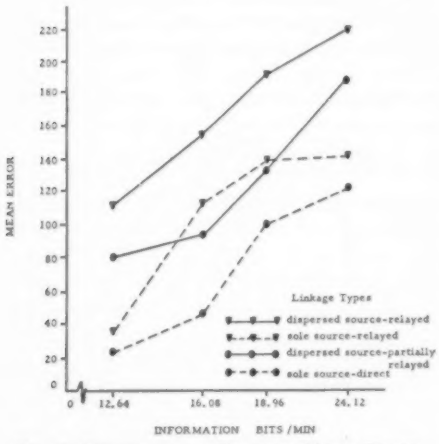

FIG. 2. RELATIONSHIPS BETWEEN MEAN ERROR SCORES AND INPUT RATE FOR THE FOUR LINKAGE TYPES

significant. The interaction between linkage types and "uncertainty" is not significant.

The significance of "sole" vs. "dispersed" source indicates why the relative superiority of HA over LA does not increase as the uncertainty of the input to the group increases. Errors for sole-source controls appear to rise rapidly at first and then level off, whereas errors for dispersed-source controls continue to increase linearly as uncertainty increases (see Fig. 2). Since both LA and HA include sole-source controls as well as dispersed-source controls, mean errors for these structures tend to exhibit similar relationships with "uncertainty."

In terms of decreasing total error scores, the linkage types order themselves as follows: sole source—direct; sole source—relayed; dispersed source—partially relayed; and dispersed source—totally relayed. Some rather simple assumptions may explain this ordering. One may postulate that correct control actions are dependent on one or more of three primary processes—detection of a change in an instrument reading, communication of that change to the individual requiring the information, and translation of the information into control actions by reference to a table of operating instructions. The latter activity is common to all the controls but the two other processes are required to a differential degree according to linkage type.

Assuming that there is a certain probability of correct "detection" and "communication" and that such probabilities are independent across operators, then the probability of a correct setting for dispersed source linkages is a function of the joint probability of detection by two operators, while the probability of a correct setting for sole source linkages is a function of the probability of detection by a single operator. Similarly, linkages requiring the communication of information from two operators are dependent on the joint probability of correct communication from two operators while those requiring only a single communication are a function of the probability of correct communication from one operator. Assuming further that for all operators the probability of "detection" and "communication" are approximately the same, the joint probabilities of detection or communication reduce to the square of the probability for a single operator. Designating the probability

of detection by P_d and of communication by P_c, the probability of correct control actions for the four linkage types would be as follows:

1. Sole source—direct:

$$P_1 = f(P_d).$$

2. Sole source—relayed:

$$P_2 = f(P_d \cdot P_c).$$

3. Dispersed source—partially relayed:

$$P_3 = f(P_d^2 \cdot P_c).$$

4. Dispersed source—totally relayed:

$$P_4 = f(P_d^2 \cdot P_c^2).$$

It is apparent that for all values of P_d and P_c except 1, P_d^2 and P_c^2 will be less than P_d and P_c respectively, and thus that the linkage types would order themselves as was empirically found. The probability analysis, of course, implies much more than a simple ordering, but a quantitative test of the deductions must await the collection of data which would allow an estimation of the parameters. In the present data there is no way of accurately separating errors of detection from errors of communication or translation.

SUMMARY AND CONCLUSIONS

Three-man groups performed a simple "team" task involving the processing of presented instrument readings, the relaying of necessary information to individuals requiring it, and the execution of control adjustments based on relayed or directly available instrument readings. The loci of controls were varied to produce two work structures differing in the extent to which they required the transmission of information. In addition, the rate of change of the instrument readings, the predictability of the changes among individuals, and the sequence of instrument changes for an individual, were manipulated. The three task variables were arranged in a $2 \times 2 \times 2$ factorial design with two groups assigned to each cell. Each of the two groups performed under both structure conditions, the order of presentation of the conditions being counterbalanced for each cell. The performance measure used was the total number of times each control was incorrectly adjusted.

In agreement with results previously obtained (7), the most difficult structure was that in which a larger proportion of information had to be relayed and, more critically, in which a larger proportion of information had to be relayed from several different sources. The importance of the latter factor was reinforced by the analysis of linkage types (in terms of whether the individual's sources of information were direct or relayed, and concentrated in a single source or dispersed) which indicated that the results for structure may be compounded of a specific linkage-type effect and a general context or difficulty effect, the latter being primarily a function of the volume of information to be relayed, the former primarily reflecting the extent of dispersion of relevant information.

Errors for both structure conditions significantly increased when the rate of change of instrument readings increased. There were no significant effects attributable to an increase in the sequential dependency between instrument changes, presumably because such dependencies were not fully perceived. An information measure applied to the task variables indicated an increasing monotonic relationship between errors and information rate. There were a considerable number of errors even at the lowest rate of approximately 12 bits per minute, a rate considerably lower than the estimates of channel capacity obtained with individuals in discrimination and reaction-time studies (e.g., 5).

This finding and the results for structure would appear to indicate that the limiting factor in the performance of the groups was not their gross information capacity. Rather the difficulty seems to lie in the inability of groups to set up an efficient system for detecting and communicating information changes. Communication problems may result from ignorance on the part of response agents as to when information bearing on their controls enters the group at some other station, and on the part of information-source persons as to the relevance of new information they receive. Detection difficulties may be a function of a response conflict generated by placing the individual in the dual role of response agent and information source. At any rate, the assumption of a high degree of dependence of control actions on both "detection" and "communication," with some further assump-

tions concerning the theoretical reliability of these processes for the various linkage types, accounts for the ordering of linkage types obtained. The interpretation is only weakly supported by the results, however, since the probability analysis implies much more than a simple ordering. Communication records are currently being obtained from which more decisive evidence may be gleaned.

REFERENCES

1. ANDREAS, B. G., & WEISS, B. *Review of research on perceptual-motor performance under varied display control relationships.* Rochester: Univer. of Rochester, May 1954. (Contract No. AF 30(602)-200, Sci. Rep. No. 2.)
2. EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
3. GROSSMAN, E. R. Entropy and choice time: the effect of frequency unbalanced on choice response. *Brit. Quart. exp. Psychol.*, 1953, **5**, 41-52.
4. HYMAN, R. Stimulus information as a determinant of reaction times. *J. exp. Psychol.*, 1953, **45**, 188-196.
5. LUCE, R. D. *A survey of the theory of selective information and some of its behavioral applications.* New York: Columbia Univer., Bureau of Applied Social Research, June 1954. (Tech. Rep. No. 8.)
6. MCMILLAN, B., GRANT, D. A., *et al. Current trends in information theory.* Pittsburgh: Univer. of Pittsburgh Press, 1954.
7. ROBY, T. B., & LANZETTA, J. T. An investigation of task performance as a function of certain aspects of work-group structure. *USAF Personnel Train. Res. Cent. Res. Rep.* submitted for publication. (Unpublished manuscript on file, Crew Research Laboratory, Randolph Air Force Base, Texas.)
8. ROBY, T. B., & LANZETTA, J. T. Work group structure, communications, and group performance. *Sociometry*, in press.
9. SHAW, M. E. Some effects of unequal distribution of information upon group performance in various communication nets. *J. abnorm. soc. Psychol.*, 1954, **49**, 547-553.

Received December 27, 1955.

EFFECTS OF GLUTAMIC ACID AND SOCIAL STIMULATION IN MENTAL DEFICIENCY

GUINEVERE S. CHAMBERS AND RALPH N. ZABARENKO[1]

Western Psychiatric Institute, University of Pittsburgh[2]

INTEREST in the effects of glutamic acid on mental functioning was aroused by a publication in 1943 (10) reporting observations made in conjunction with the use of the drug as a possible treatment in convulsive disorders. Subsequent studies on its effects on learning ability in animals (2, 23) led to extensive investigations with humans (1, 14, 18, 19, 20, 21, 22). A review of the literature reveals that interest in this area of research reached a peak in 1951–52. Since then, there has been an abrupt falling off of such publications in this country with a concomitant growth of interest in the problem in other countries.

Gadson, reviewing the literature to 1951 gives a concise presentation of the animal and human studies that yielded positive results. At the same time there have been a number of investigations, both animal (7, 9, 17) and human (4, 8, 11), where negative or doubtful results have been obtained.

With research on the wane in this area, the picture remains confused. It is difficult to compare results from one animal study to another because of basic differences in research design, types of learning being tested, number of strains of animals and finally, the difference in form and quantity of glutamic acid used. Studies with humans are even more confusing when compared. In the present study an attempt was made to overcome some of the weaknesses in other experiments by:

1. Using a large sample of institutionalized defectives representing a narrow range of intellectual level.

2. Maintaining 24-hour-a-day supervision of their environment and medication.

3. Controlling accuracy of medication by the use of tablets administered by trained personnel.[3]

4. Using comparable tests throughout the experiment and avoiding the introduction of nonexperimental variables.[4]

5. Including a placebo to run concurrently with the glutamic acid therapy.

6. Avoiding a "halo" effect by not revealing which tablet is the drug and which the placebo.

7. Describing in detail the environmental surroundings which conceivably could have had a bearing on results.

METHOD

Subjects. The 50 male and 8 female *S*s used in the experiment were residents of Polk State School, an institution for the care and training of mental defectives. On selection, IQs ranged from 20 to 66 with a mean of 54. To permit evaluation in terms of differential effects which might be attributable to differences in mental growth rates, two distinct age groups were represented. An immature group consisting of 32 *S*s ranged in age from 5 to 11. The remaining 26 *S*s constituted a mature group with an age range of 19 to 35. Previous testing at the institution, in many cases covering a ten-year period, indicated that the IQs at this intelligence level were quite stable.

In the hope of isolating differential reaction to the drug, *S*s were selected from three etiological categories: (*a*) familial mental defect; (*b*) defect due to abnormal developmental conditions; (*c*) postnatal brain damage. Selection of *S*s was aimed at avoiding cases with psychoses or convulsive disorders. However, during the course of the experiment one mature *S* gave evidence of bizarre ideation and 2 immature *S*s displayed convulsive behavior, but all were retained throughout the study. No mongoloids or phenylpyruvic oligophrenics were included.

Experimental environments. The *S*s were treated in two distinctly different environments in an effort to answer the question asked of other studies in which increase in intellectual functioning is under scrutiny: How much noted change is the result of the experimental variable and how much to an increase in social opportunities and a change in self concept?.

[1] Formerly Associate Research Psychiatrist, Western Psychiatric Institute, Pittsburgh, and presently, Instructor, University of Pittsburgh School of Medicine.

[2] The authors wish to express their appreciation to: Dr. G. B. Pearson, Former Director of the Institute; Dr. Gale H. Walker, Superintendent of Polk State School; Dr. R. A. Patton, Research Coordinator for the project; Dr. A. W. Bendig, who gave statistical advice; Mrs. Frances Williams, Former Directress of Nurses; Mrs. Helen Galloway and Miss Christine Haines, Dieticians; and especially to nurses, attendants, and other personnel too numerous to be given personal acknowledgment.

[3] Zimmerman *et al.* (23) comment on the difficulty of getting parents to administer the drug faithfully.

[4] *Ibid.* These workers include cases who had been receiving psychotherapy and others who were being intensively tutored.

Twenty-seven male *Ss* (16 immature, 11 mature) were transferred to the Western Psychiatric Institute and Clinic in Pittsburgh (WPI). The other 31 *Ss* remained in their accustomed environment at Polk. All *Ss* were selected by the medical director of the project from the imbecile and low-grade moron groups for comparability of institutional history and similarity of defect within each subgroup, and were randomly assigned to the WPI and Polk groups. It developed that there was a significant difference in mean IQ between these two groups (44.09 and 34.59 for the Polk and the WPI groups respectively).[5]

Polk School, a well-run institution, is located in a rural area that allows considerable freedom of movement for patients. Turnover in personnel is low. In many instances attendants are third generation employees who have taken on family roles in relation to the patients. Assignment of living quarters in homelike cottages is made according to age, sex, and degree of defect. Throughout the study, the living conditions of the Polk group remained essentially the same except for those of eight small boys. To simplify the medication procedures, these *Ss* were transferred to a unit which they shared with approximately 90 adult males. This resulted in their receiving almost constant attention and becoming cottage favorites. All medication was administered by two graduate psychiatric nurses from WPI who resided at Polk for the duration of the experiment. In addition to the medicating, they made detailed behavioral notes but were not responsible for the routine care of the *Ss*.

In sharp contrast to Polk, WPI, a modern structure with hospital atmosphere, is located in the business and cultural center of a city. One floor was set aside as living quarters for the *Ss*. The ratio of personnel to patients was much higher in the WPI situation than at Polk. Freedom to initiate and carry out individual activities was greatly restricted at WPI because of its urban setting. The city environment, on the other hand, provided opportunity for many new experiences. Relatives made more frequent visits to patients because of the better transportation facilities in the urban area, and also, apparently, because they had renewed interests and hopes for these children about whom they had previously adopted an attitude of resignation.

Experimental schedule. Because of the added variable of change in environment for the WPI group, psychological tests were administered on five occasions in contrast to three for the Polk group. These additional testings occurred in the initial phase of the experiment to ascertain the effects of change of environment before the introduction of medication. The WPI group was tested (*a*) on selection; (*b*) immediately after transfer to the Institute; (*c*) at the end of a six-month adaptation period; (*d*) at the close of the first medication period; (*e*) at the end of the second medication period. Testing for the Polk group corresponds to administrations *a*, *d*, and *e* for the WPI group. Both at Polk and WPI the mature and immature groups were divided into two treatment subgroups. Each *S* served as his own control by being in the drug group for one treatment period and in the placebo group for the other. Glutamic acid and placebo treatment ran concurrently throughout.

Medication. L+ glutamic acid made up into 0.5-gram tablets was used as the experimental drug, while a placebo consisting of starch, lactose, citric acid, and oil of peppermint was the control treatment. The drug and placebo were identical in appearance and very similar in taste. Except for the medical director, none of the personnel knew which medication contained the glutamic acid, although it was known that a placebo was being used. Psychological test results were in the possession of the advisory committee before the psychologist knew when each group had been in the experimental or control phase.

In the early stages of the experiment, the dosage levels were increased very gradually, but when no disturbing effects were noted at maximal dosage, later groups were brought to full 40-gram intake over a period of 10 days. Each *S* was given daily doses of 40 grams of glutamic acid or an equivalent number of tablets of the placebo during the control period. Medication was administered at the same hours each day, three times a day. The bulk of the daily dose was given early in the day to avoid possible evening restlessness, e.g., at maximal dosage each group was receiving 39 tablets at 10 A.M., 27 at 3 P.M., and 14 at 8 P.M. The medication periods were at least 100 days in length, with somewhat longer periods for several groups necessitated by the amount of time required for travel and testing. Each *S* was individually medicated. In effecting the change-overs in type of medication, the total number of tablets remained the same, a few tablets of the new medication being gradually substituted for the old until full dosage was reached at the end of a 10-day period.

With rare exception, the *Ss* were entirely cooperative in taking the medication. The mature *Ss* were especially eager for their treatment, feeling that as a result they were going to be "stronger" and "smarter." The fact that these *Ss* recognized themselves as a small segment of the Polk population in itself lent status to their position. In the approximately 10,000 patient days of medication there were at most a half dozen instances of gastric disturbances. A relationship to medication could not be demonstrated.

Testing procedures. All psychological tests were administered by one psychologist who became known to the *Ss* before initiating testing to insure adequate communication and maximal cooperation.

The test battery for the mature *Ss* was: (*a*) Stanford-Binet, Forms L and M alternated; (*b*) Grace Arthur Point Scale of Performance Tests; (*c*) Rorschach; (*d*) Vineland Social Maturity Scale; (*e*) Bender-Gestalt; (*f*) Human Figure Drawings; (*g*) Test of Animistic Thinking (12); (*h*) Test of Concept Formation; (*i*) Minnesota Rate of Manipulation. The battery for the immature groups was (*a*) Stanford-Binet or Merrill-Palmer;[6] (*b*) Vineland Social Maturity Scale.

RESULTS

Psychiatric observations. A report of the psychiatric aspects of the study is presented in

[5] More clear-cut conclusions could have been reached in regard to the two environments had this difference been anticipated.

[6] Mental ages for 4 immature *Ss* are from the Merrill-Palmer Scale because of inadequate language development.

the initial report on this research (16). Briefly, the most pertinent findings were that the *Ss* who were removed from their stable, familiar surroundings showed marked signs of distress both in physiological functioning and regressive social behavior, which quickly disappeared. Not only did *Ss* in this group regain former competence, but they also gave indications of response to the added attention by appearing more alert and developing more complicated patterns of social behavior. There was a low ceiling to the amount of social growth that could be stimulated, and they remained on this plateau for many months. Toward the end of the study they found the life at WPI restrictive and lost some of their earlier hope of becoming "smarter."

The positive changes observed in the WPI group were also seen in the Polk group, but these continued to make their appearance over a longer period than at WPI. These changes were especially notable in the group of eight small boys who had been transferred to the adult cottage.

Psychological test results. Considering first the effect of glutamic acid as measured by intelligence test scores, analysis of variance revealed no significant differences in the means for any group between the drug and placebo treatments. The data are given in Table 1. An analysis was made of the Stanford-Binet data for *Ss* who gained 10 or more months in mental age during the treatment period. The cumulated gains were broken down into those achieved when the drug was the first medication and those occurring when the placebo was administered first. The largest gain made by any one individual during the treatment period was 19 months. The difference between the drug and placebo gains again was not significant nor related to the order of medication.

TABLE 1
COMPARISON OF PERFORMANCE IN MENTAL AGE MONTHS OF THE POLK AND WPI GROUPS UNDER GLUTAMIC ACID AND PLACEBO TREATMENT

Source of Variation	*df*	Mean Square	*F* Ratio
Between Groups	3	6591.887	19.570**
Ss in Same Group	54	336.844	
Between Treatments	1	42.241	2.134
Interaction: Treatment × Groups	3	2.346	<1.0
Interaction: Pooled *Ss* × Treatment	54	19.741	
Total	115		

** $p < .001$.

TABLE 2
COMPARISON OF PERFORMANCE IN MENTAL AGE MONTHS OF THE WPI GROUP TESTED UNDER THREE ENVIRONMENTAL CONDITIONS

Source of Variation	*df*	Mean Square	*F* Ratio
Between Maturity Levels	1	9508.94	23.73**
Between *Ss* in Same Group	25	400.69	
Between Conditions	2	128.59	12.57*
Interaction: Conditions × Maturity Level	2	8.95	<1.0
Interaction: Pooled *Ss* × Conditions	50	10.23	
Total	80		

* $p < .01$.
** $p < .001$.

The data from the first three testings on the Stanford-Binet for the WPI group were examined to assess the effects of transfer and an adaption period. The differences in these means were significant (.01 level, Table 2). Applying the *t* test, the source of the difference could be attributed to gains accrued during the socialization period, between transfer (M = 48.5) and the end of the six months of adaptation (M = 51.33), the mature and immature *Ss* having contributed equally to the total gain which was significant at the .001 level. The difference between initial testing performance (M = 47.0) and that following transfer was not significant.

Participation in the experiment was accompanied by significant differences in Stanford-Binet scores in all groups (.001 level) even though there was no difference found between the medication and placebo effects. In consideration of the 16-month time lapse and the inclusion of immature *Ss* who constitute over half the total number, and who would be expected to be developing, even at a retarded rate, the over-all change was not impressive.

Considering total groups, immature *Ss* made greater gains than mature ones (.05 level), and the Polk group made greater mean gains than the WPI group (.01 level). Since there was a significant difference between the WPI and Polk groups at the time of selection, a further analysis was made in an effort to determine whether ability to respond to environ-

TABLE 3
MONTHS MENTAL AGE CHANGE OF EIGHT IMMATURE MALE SUBJECTS

Case Number	Glutamic Acid	Placebo	Over-all Gain
43	5	1	6
46	0	6	6
47	1	3	4
50	2	13	15
51	13	1	14
53	8	3	11
54	17	1	18
57	5	10	15

mental stimulation may be dependent upon a minimal basic capacity. All *Ss* were dichotomized into those above and below IQ 45. Again, the relationship held for maturity level, with immature *Ss* making significantly greater gains (.05 level) than mature *Ss* with the younger *Ss* having IQs above 45 making the greatest gains.

The changes in mental age scores of the eight small boys at Polk (Table 3) are of especial interest since, with one exception, they maintained an approximately normal rate of gain in mental age over the experimental period.

While there were no differential reactions to the drug among the three diagnostic categories, there were significant differences between groups (.05 level) in terms of participation in the study. The postnatal defect group (holding location constant) made significantly greater gains than did the familial or developmental defect groups.

Analysis of the mean Grace-Arthur scores revealed no significant differences in performance for any group or aspect of the study. This is rather surprising in that the tasks involved would seem to be more susceptible to practice effect than would the verbal material.

There were changes in the Rorschach performances which are reported elsewhere (3). These changes could not be attributed to the administration of glutamic acid. Changes in the Human Figure Drawings closely paralleled those noted in the Rorschach and appeared to be related to greater familiarity with the task or possibly some enhancement in self concept.

No changes were demonstrable on any other tests. Even though *Ss* did develop new skills such as bed-making and self-care at the table, these were at a sufficiently uncomplicated level that they made little change in Vineland Quotients. Most of the gain occurred in areas where formerly there had been no opportunity to develop these skills.

DISCUSSION AND CONCLUSIONS

Since no tests applied to the data yielded significance between drug and placebo medication, it must be concluded that for these *Ss* glutamic acid did not improve mental functioning. For many *Ss* the test scores following the glutamic acid treatment were numerically the lowest. It was the examiner's opinion, based on behavioral notes made at each testing session, that under the drug therapy there was some lessening in ability to inhibit responses which led to impulsive errors, especially on the motor tasks.

It is of interest that the mental-age scores of these patients did not change immediately after transfer when their environment was drastically altered. The most plausible explanation for this appears to be that after an initial reaction to change (nausea, sleeplessness, crying, and loss of previous skills in self-care), they quickly returned to long established patterns of response. The finding does not appear to be attributable to insensitivity in the clinical instruments used, since mental-age scores that had been remarkably stable over many years did change significantly after a treatment-free adaptation period. Studies on the effects of glutamic acid which did not include this control period could have reported significant gains which were the product of the uncontrolled variable—increased attention—rather than the effect of the drug. It cannot be overemphasized that any environmental manipulation of a mental defective is an important variable in an experiment.

The research was designed to keep the Polk environment as similar to pre-experimental conditions as possible. Yet, the awareness on the part of the attendants that these *Ss* were participating in the experiment actually produced considerable change in this environment. As these people were being periodically queried about changes noted in the *Ss*, they became more aware of the patients as individuals and were quite hopeful that their particular charges would be improved with the drug. While conclusions from this research on institutionalized *Ss* are not directly applicable to studies of non-

institutionalized defectives, it is likely that in those situations, too, the *S*s were receiving added attention and were responding to new hopes held for them by parents. Sarason's (13) survey of the literature on the effects of environmental stimulation adds support to this hypothesis.

Because more of the Polk *S*s initially were at the higher end of this narrow intellectual range, the differential effects of the two environments are somewhat clouded. The important fact is that *both* groups received added attention and both groups made over-all gains, not related to medication. The finding of the greater gain having occurred at Polk was anticipated by directing personnel. Individuals accustomed to handling mental defectives have long been aware of the fostering effects of uninterrupted routine on their comfort and level of performance. Although the WPI *S*s obviously enjoyed some novel experience such as a trip to the zoo, the changes in scheduling necessitated to provide these activities always carried with them considerable restlessness and distress. Frequent changes in personnel at WPI also seemed to make this a less secure environment than was offered at Polk.

It is not surprising that the younger children were more responsive to the stimulation provided by the experiment than were the adults. It seems logical that if a spurt of development can be induced, it is more likely to occur in an organism in the process of developing than in one where an apparent ceiling has been previously attained. When the scores of the mature *S*s from childhood through their institutional life were plotted, a peak was reached typically at age 10 to 12 followed by a gradual decline to a long-standing plateau. The changes that occurred during the experiment did not exceed the highest scores on early tests, so that rather than representing increased capacity, they seemed merely to reflect higher motivation to utilize native potential. If such was the case, differentials in gains may possibly be a function of the strength of individual desire to improve.

While it is impossible from these data to account for the marked differences in over-all reaction to the experiment, the most promising lead comes from consideration of the group of boys at Polk who had, prior to the experiment, been developing at the rate of 25 to 50 per cent of normal expectation. It was within this subgroup that the greatest and most consistent gains occurred. These children had an especially enriched environment for the duration of the study. Possibly awareness of the concern of others coupled with "tender loving care" may account for over-all gains reported in this study as well as in other studies where positive findings on small samples have been attributed to glutamic acid.

SUMMARY

1. L+ glutamic acid in this study did not produce significant improvement in mental functioning.

2. Abruptly changing the environment of institutionalized mental defectives did not affect mean mental-age scores as measured by the Stanford-Binet Intelligence Scale and Grace - Arthur Point Scale of Performance Tests.

3. Increased personal attention enhanced mental functioning.

4. A significantly greater gain was made in functioning of mental defectives who were given added attention but with minimal disruption of the familiar environment.

5. Immature *S*s with IQs of 45 or above in this study made the most dramatic changes in mental functioning.

6. Subjects diagnosed as being defective as the result of postnatal insult made significantly greater gains than did *S*s diagnosed as being familial defectives or having had abnormal prenatal development.

REFERENCES

1. ALBERT, K., HOCH, P., & WAELSCH, H. Preliminary report on the effect of glutamic acid administration in mentally retarded subjects. *J. nerv. ment. Dis.*, 1946, **104**, 263–274.
2. ALBERT, K. E., & WARDEN, C. J. Level of performance in the white rat. *Science*, 1944, **100**, 476.
3. CHAMBERS, G., & HAMLIN, R. Rorschach "inner life" capacity of imbeciles under varied conditions. *Amer. J. ment. Def.*, in press.
4. ELLSON, D. G., FULLER, P. R., & URMSTON, R. The influence of glutamic acid on test performance. *Science*, 1951, **112**, 248–250.
5. GADSON, E. J. Glutamic acid and mental deficiency —A review. *Amer. J. ment. Def.*, 1951, **55**, 521–528.
6. HARNEY, SISTER MAUREEN. *Some psychological and physical characteristics of retarded girls before and following treatment with glutamic acid.* Washington, D. C.: Catholic Univer. of America Press, 1950.

7. MARX, M. H. Effects of supranormal glutamic acid on maze learning. *J. comp. physiol. Psychol.*, 1948, **41**, 82–92.
8. MCCULLOCH, T. L. The effect of glutamic acid feeding on cognitive abilities of institutionalized mental defectives. *Amer. J. ment. Def.*, 1950, **55**, 117–122.
9. PILGRIM, F. J., ZABARENKO, L. M., & PATTON, R. A. The role of amino acid supplementation and dietary protein level in serial learning performance of rats. *J. comp. Physiol.*, 1951, **44**, 26–36.
10. PRICE, J. C., WAELSCH, H., & PUTNAM, T. J. Glutamic acid hydrochloride in the treatment of petit mal and psychomotor seizures. *J. Amer. Med. Ass.*, 1943, **122**, 1153–1156.
11. QUINN, K. V., & DURLING, D. New experiment in glutamic acid therapy; 24 cases classified as mental deficiency, undifferentiated, treated with glutamic acid for 6 months. *Amer. J. ment. Def.*, 1950, **55**, 227–234.
12. RUSSELL, R. W., & DENNIS, W. Studies in Animism: I A standardized procedure for the investigation of animism. *J. genet. Psychol.*, 1939, **55**, 389–400.
13. SARASON, S. B. Psychological problems in mental deficiency. New York: Harper, 1949.
14. WAELSCH, H. A biochemical consideration of mental deficiency. The role of glutamic acid. *Amer. J. ment. Def.*, 1948, **52**, 305–313.
15. ZABARENKO, L. M., PILGRIM, F. J., & PATTON, R. A. The effect of glutamic acid supplemental on problem solving of the instrumental conditioning types. *J. comp. physiol. Psychol.*, 1951, **44**, 126–133.
16. ZABARENKO, R. N., & CHAMBERS, G. S. An evaluation of glutamic acid in mental deficiency. *Amer. J. Psychiat.*, 1952, **108**, 881–887.
17. ZIMMERMAN, F. T. Glutamic acid treatment of mental retardation. *Quart. Rev. Psychiat. Neurol.*, 1949, **4**, 263–269.
18. ZIMMERMAN, F. T., BURGEMEISTER, B. B., & PUTNAM, T. J. A group study of the effect of glutamic acid upon the mental functioning in children and adolescents. *Psychosom. Med.*, 1947, **9**, 175–183.
19. ZIMMERMAN, F. T., BURGEMEISTER, B. B., & PUTNAM, T. J. Effect of glutamic acid on mental functioning in children and in adolescents. *Arch. Neurol. Psychiat.*, 1946, **56**, 489–506.
20. ZIMMERMAN, F. T., BURGEMEISTER, B. B., & PUTNAM, T. J., Effect of glutamic acid on the intelligence of patients with Mongolism. *Arch. Neurol. Psychiat.*, 1949, **61**, 275–287.
21. ZIMMERMAN, F. T., BURGEMEISTER, B. B., & PUTNAM, T. J., The ceiling effect of glutamic acid upon intelligence in children and in adolescents. *Amer. J. Psychiat.*, 1948, **104**, 593–599.
22. ZIMMERMAN, F. T., BURGEMEISTER, B. B., & PUTNAM, T. J. The effect of glutamic acid upon the mental and physical growth of Mongols. *Amer. J. Psychiat.*, 1949, **105**, 661–668.
23. ZIMMERMAN, F., & ROSS, S. Effect of glutamic acid and other amino acids on maze learning in the white rat. *Arch. Neurol. Psychiat.*, 1944, **51**, 446–451.

Received September 21, 1955.

RETENTION OF AFFECTIVELY TONED VERBAL MATERIAL BY NORMALS AND NEUROTICS[1]

SAMUEL F. KLUGMAN

VA Hospital, Coatesville, Pennsylvania

FOLLOWING Freud's attempt to relate forgetting to repression, a relatively large number of experiments have been carried out to study this relationship. An examination of these studies reveals many ambiguous results and weaknesses in procedure which have been ably summarized (1, 5, 11, 13, 19).

In this study, an attempt has been made to overcome the following methodological weaknesses found in earlier investigations: failure to equate for familiarity of the verbal material to be retained, insufficient number of *S*s, inadequate control of overlearning and of early mastery of the learned material, neglect of intensity factors, and use of isolated words to convey emotional ideas. Furthermore, all of the quantitative studies, except Sharp's (17) which supported and Kott's (9) which failed to support Freud's theory of repression, dealt only with normal *S*s. If emotional factors have a significant effect upon memory, then certainly the emotional stability of the *S* should be considered. In addition, "Each *S* must be treated with individual attention to his particular complexes . . ." (16, p. 119).

Freud's theory of repression and its relation to conversion symptoms may be related to levels of ego strength as found in normals and neurotics in such a way as to permit the prediction of differential retention between these groups in terms of quality (pleasantness-unpleasantness) and intensity (strong-mild) variables. According to Freud (3), an hysterical conversion symptom is due to a displaced affect that was part of an original traumatic incident, the recall of which would be painful to the individual. The ideational part of the original incident is subject to repression, a process serving "*the function of rejecting and keeping something out of consciousness*" (p. 86, italics his). At a later point he states that, for neurotic patients, repression makes it "possible to bring about a total disappearance of the charge of affect" (p. 94). For this reason, in the present study, only those neurotics were used who gave evidence of conversion symptoms. It was felt that a population of this type, if Freud is correct, would yield the most definitive results since the *S*s may be conceived as typically employing repression as a defense mechanism.[2]

Hypotheses

The general hypothesis of this study is that normal and neurotic *S*s are differentially affected in their retention of learned verbal material by its affective quality and its intensity. Verbal material was employed because, according to Freud, "repression denies to the rejected idea . . . translation of the idea into words . . . . The idea which is not put into words . . . then remains in the unconscious in a state of repression" (3, p. 134).

Freud further argued that, "From any operation which might arouse unpleasantness ('pain') mental activity draws back (repression)" (3, p. 14), and "The ego pursues pleasure and seeks to avoid unpleasure" (4, p. 16). In

[1] This paper is based on a dissertation submitted in partial fulfillment of the requirements of the degree of Doctor of Philosophy, in the Department of Psychology of the University of Pennsylvania. The writer wishes to express his sincere appreciation to the dissertation adviser, Dr. Julius Wishner, and to the other members of the committee, Dr. Francis W. Irwin and Dr. Miles S. Murphy, for the invaluable assistance so generously given throughout the experimental investigation.

[2] One effect of this procedure was to restrict the number of neurotics eligible for this experiment, making it impracticable to match the two subgroups. According to Edwards (2, p. 355), analysis of covariance is "applicable to those situations where the matching of groups is not feasible prior to the assignment of the *S*s of the experimental conditions, but where some measure of initial performance may be obtained after the assignment." Furthermore, as Lindquist points out, even if the means of the initial scores for the treatment groups are not different, analysis of covariance is desirable. "Assuming some correlation between X and Y, the within-groups variance of the adjusted measures would be less than that of the unadjusted measures, so that the precision of the experiment would be increased" (10, p. 325). Following Kott (9), it was expected that a high relationship existed between learning and retention. For these reasons, analysis of covariance was used in the experiment, and both measures of retention, recall and relearning, were adjusted for initial learning scores.

terms of retention, one can derive the following hypotheses:

1. Both normals and neurotics retain more pleasant than unpleasant material.

Freud's first statement, however, also implies that the weaker ego of the neurotic resorts more readily to repression when dealing with personally meaningful unpleasant material than does the stronger ego of the normal. It follows that,

2. Normals retain more personally relevant unpleasant material than do neurotics.

In his discussion of repression, Freud has also stated, "As soon as an idea which is fundamentally offensive exceeds a certain degree of strength, the conflict (with conscious control) takes on actuality and it is precisely activation of the idea that leads to its repression" (3, p. 90). Although it may be inferred that Freud was referring only to unpleasant ideas, any strongly intense material may be generally regarded by the neurotic as "offensive." Clinical observation and Rorschach color theory (8, p. 390) support the contention that neurotics cannot properly handle affective situations and, therefore, tend to avoid highly emotional involvement. One may then hypothesize that,

3. Neurotics retain more mildly toned than strongly toned material.

Since the weaker ego of the neurotic evokes a greater need to avoid the threat of the possible appearance of a "fundamentally offensive" idea into consciousness than does the ego of the normal, it may be finally hypothesized that,

4. Normals retain more personally relevant strongly toned material than do neurotics.

METHOD

Experimental Design

Sixty male normals, World War II veterans (never treated psychiatrically), were compared with sixty male neurotics (currently being treated at a Veterans Administration Mental Hygiene Clinic) for the four possible combinations of emotional intensity (strong-mild) and quality (pleasant-unpleasant) values. Thus, *Ss* in each group were randomly assigned to one of four subgroups. Two subgroups, one of them neurotic and one normal, learned, recalled, and relearned a set of strongly pleasant phrases; two other subgroups did the same with mildly pleasant phrases; two other subgroups dealt with mildly unpleasant phrases; and two with strongly unpleasant phrases. The arrangement of the obtained data in Table 1 illustrates the design of the experiment.

Subjects

The neurotics were selected by their therapists, who were instructed to choose *Ss* who had attended high school, were between the ages of 25 and 45, and had some conversion symptom. The normals were obtained from veterans' schools, a neighborhood men's club, and a fraternal lodge in such a way as to meet the same age and educational standards as those of the neurotics. The mean age for the normals was 30.70, *SD* 4.37; for the neurotics, the mean was 29.77, *SD* 3.52. The mean educational level for the neurotics was 11.45 grades, *SD* 2.56; and 11.02, *SD* 2.50, for the normals. The neurotics were diagnosed according to instructions included in VA Technical Bulletin 10A-78, entitled *Nomenclature of Psychiatric Disorders and Reactions*. There were 18 cases of anxiety reaction, 17 of conversion reaction, and 25 of somatization reaction.

The Phrase Construction Technique

The phrase construction technique was developed to obtain personally meaningful material from each *S*. A chart, made up of four pages, was used. The first page contained the directions for constructing either pleasant or unpleasant phrases, the second had a scale of feeling tone with location points for mild and strong descriptions, the third a list of 365 adjectives, and the fourth a list of 372 nouns. The words were selected from the *Teacher's Word Book of 30,000 Words* (18) by three psychologists working independently, who were instructed to select words (usually taught in Grades I through IV) that might be related to social, vocational, familial, love, and health areas. Following are some examples of phrases which were constructed by the *Ss* according to instructions to be described below: happy mother, excellent job, cold lunch, dirty home, cruel father, faithful wife, happy children, and crying baby. This plan provided for (*a*) permitting *S* to select materials personally meaningful to him, (*b*) equating for familiarity and difficulty, and (*c*) use of more than isolated words in an attempt to obtain more meaningful ideational content. Thus, Sears' concern that "each subject must be treated with individual attention to his particular complexes" (16, p. 119) appears to have been met to some extent.

The degree of consistency in the ratings of the phrases for emotional intensity during the one-week interval was determined. Tetrachoric correlation coefficients ranged from .77 to .84 for consistency in rating for the group as a whole and for the separate subgroups. Reliability seemed sufficient to justify further study of the experimental phrases.

Learning and Retention Procedures

First session. The following description of the first normal *S*, who worked with mildly unpleasant material, is illustrative of the experimental method. He was told to construct 24 phrases which he considered unpleasant according to the directions contained in the chart, half of them of strong and half of mild intensity. Since he was to work with mildly unpleasant material, only his mild phrases were utilized for the experiment proper.[3]

[3] Similar approaches were taken by *Ss* working with strongly unpleasant, mildly pleasant, and strongly

While *E* then printed by hand a set of learning cards, *S* read some irrelevant material[4] which was intended to minimize any rehearsal effects. Following the printing of the learning cards and the preparation of the recording form, *S* was requested to learn the phrases, each of which was associated with a nonsense syllable. The learning procedure was carried out by a combination of the paired associate and prompting methods. The same nonsense syllables, all of which had a zero association value according to Glaze's (6) classification, were used for all *S*s.

The phrases were learned to the criterion of two successive perfect repetitions in the following manner: first, *S* studied each association of a phrase with a nonsense syllable for ten seconds; second, the cards were shuffled and placed in an exposure box (with an opening to the left so that only the syllable of the top card was exposed); third, *S* tried to recall the phrase in five seconds. If he recalled incorrectly or not at all within that time, he was allowed to study the card for five seconds more. After each trial (consisting of going through the set of cards once), the cards were shuffled in a manner suggested by Osgood (12) to prevent serial learning. As soon as *S* had learned a card to the criterion, it was removed from the set to prevent overlearning. After every fifth trial and at the conclusion of the learning procedure, *S* read once aloud each of the cards that had so far been removed from the set. These review trials were given in order to equalize the degree of mastery over those items learned early with those learned more recently. The score was the number of trials necessary to learn all associations to the criterion.

In compliance with a request made by several school directors, a limit was arbitrarily set so as not to keep *S*s away from their classrooms too long. If *S* took more than twenty trials to learn, he was removed from the experiment and the next *S* took his place.

Between this session and the next, *E* typed *S*'s original phrases, arranged in random order, on a mimeographed form containing directions for reclassification of the phrases at the conclusion of the next session.

Second session. During the second session, which occurred exactly one week later, at the same hour, *S* was required to recall as many phrases as he could from a set of cards that contained only the nonsense syllables. He was allowed 30 seconds in which to recall the associated phrase. His score was the number of words correctly recalled. Following this, his original set of learning cards was placed in the exposure box, and *S* was asked to relearn the associations. Except for the fact that he did not study each card initially for ten seconds, the procedure was exactly the same for relearning as for learning. Finally, *S* was presented with the reclassification form containing his 24 original phrases, randomly arranged, and directed to indicate his present attitude as to whether each phrase was mild or strong in emotional intensity.

RESULTS

Learning

The mean learning scores reported in Table 1 represent the number of trials necessary to learn the phrases to the point of two consecutive perfect repetitions. It is apparent that the two groups, in general, were alike in their ability to learn this type of material. An analysis of variance of these scores indicated no significant differences among the variables singly or in interaction. In the case of the triple interaction, a significant inverse *F* ratio at the .05 level was obtained, but as Edwards (2, p. 228) points out, such a ratio has "no reasonable interpretation other than that (it is) the occasionally significant value which (is) to be expected from random sampling." Bartlett's test for homogeneity of variance was nonsignificant, and inspection of the distributions in the separate cells showed no unusual skewness.

TABLE 1
LEARNING AND RETENTION MEANS FOR NEUROTIC AND NORMAL GROUPS*

	Strongly Pleasant	Mildly Pleasant	Mildly Unpleasant	Strongly Unpleasant
A. Neurotics				
Learning				
Mean	13.40	12.83	13.13	13.73
SD	2.53	2.92	3.36	2.64
Recall				
Original Mean	9.33	10.60	9.73	8.73
SD	2.10	2.06	2.28	1.71
Relearning				
Original Mean	8.40	7.40	7.73	8.67
SD	1.71	1.82	1.93	2.20
B. Normals				
Learning				
Mean	12.80	13.47	13.80	13.27
SD	2.56	2.69	2.40	2.65
Recall				
Original Mean	10.73	9.33	9.00	10.40
SD	1.79	1.64	1.67	1.78
Relearning				
Original Mean	7.53	8.53	8.00	7.27
SD	1.60	2.01	1.21	1.76

* Learning and Relearning scores: trials to learn; Recall scores: words recalled.

pleasant material. The instructions to the *S*s during the first and second sessions, and the phrase construction chart, with separate directions for pleasant and unpleasant phrase construction, have been deposited with the American Documentation Institute. Order Document No. 5033 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting $1.75 for microfilm or $2.50 for photocopies. Make check payable to Chief, Photoduplication Service, Library of Congress.

[4] *Comic Relief, a Collection of Modern American Humor*, by R. W. Linscott, New York: Houghton Mifflin, 1932.

These results, revealing no differences in learning between the groups, are similar to those found by Sharp (17), Kott (9), and Schmidt (15).

Correlations Between the Various Measures

Table 2 presents all product-moment correlations found among the various variables for each of the subgroups and the total sample. All coefficients are significant at the .01 level. In no instance, as determined by z values, was there a significant difference between any pairs of coefficients. The high correlations confirmed the need for partialling out the strong effects of learning on recall and relearning scores.

Adjusted Retention Scores

The recall data were adjusted for initial learning by analysis of covariance. As shown in Table 3, the interaction between affective intensity and normal-neurotic grouping yielded an F significant at the .01 level. No other significant Fs were found. The relearning data were similarly adjusted for learning scores. Significant Fs were obtained for the double interactions between grouping × quality and grouping × intensity.[5]

In view of these significant Fs obtained from the double interactions, tests were made to determine the significance of the differences between the groups for quality and intensity variables, respectively, and for both measures of retention. These appear in Table 4.

The upper half of the table contains the data pertinent to the hypotheses of the study. The first hypothesis stated that "both normals and neurotics retain more pleasant than unpleasant material." The results for both groups fail to support this hypothesis; in fact, the relearning data for normals indicate significantly better retention of the unpleasant than of the pleasant material. The second hypothesis stated that "normals retain more personally relevant unpleasant material than

TABLE 2

PRODUCT-MOMENT CORRELATIONS BETWEEN EXPERIMENTAL VARIABLES

Variables	Total	Neurotics	Normals
Learning vs. Recall*	−.68	−.68	−.68
Learning vs. Relearning	.86	.87	.85
Recall vs. Relearning*	−.71	−.75	−.70

* Fewer trials to learn or relearn related to greater number of words recalled.

TABLE 3

ANALYSIS OF RETENTION SCORES ADJUSTED FOR LEARNING SCORES

Source	*df*	Recall		Relearning	
		Variance	F*	Variance	F*
Pleasant-Unpleasant (P-U)	1	4.20	2.09	1.75	2.12
Strong-Mild (S-M)	1	.42	—	.14	—
Neurotic-Normal (PN-N)	1	2.00	—	1.66	2.01
P-U × S-M	1	.28	—	.01	—
P-U × PN-N	1	1.61	—	4.52	5.29
S-M × PN-N	1	29.54	14.70	10.35	12.56
P-U × S-M × PN-N	1	.15	—	.21	—
Error	111	2.01		.824	

* F .05 = 3.93; F .01 = 6.87

TABLE 4

ADJUSTED MEAN DIFFERENCES BETWEEN NORMALS AND NEUROTICS FOR QUALITY AND INTENSITY VARIABLES

	Recall			Relearning		
	Adjusted Means†		t	Adjusted Means‡		t
A. Pertinent to Hypotheses						
PN-P vs. PN-U	9.90	9.28	1.68	7.99	8.14	.64
N-P vs. N-U	9.94	9.80	.38	8.14	7.45	2.93**
N-U vs. PN-U	9.80	9.28	1.41	7.45	8.14	2.92**
PN-M vs. PN-S	10.03	9.15	2.41*	7.74	8.39	2.76**
N-S vs. PN-S	10.43	9.15	3.50**	7.57	8.39	3.49**
B. Nonpertinent						
N-S vs. N-M	10.43	9.32	3.04**	7.57	8.09	2.23*
N-S vs. PN-M	10.43	10.03	1.09	7.57	7.74	.72
PN-M vs. N-M	10.03	9.32	1.95	7.74	8.09	1.51
N-M vs. PN-S	9.32	9.15	.46	8.09	8.39	1.25
N-P vs. PN-U	9.94	9.28	1.80	8.14	8.14	.01
N-P vs. PN-P	9.94	9.90	.11	8.14	7.99	.65
PN-P vs. N-U	9.90	9.80	.27	7.99	7.45	2.28*

* Significant at .05 level of confidence.
** Significant at .01 level of confidence.
† Number of words recalled.
‡ Trials to relearn.

[5] There is some question whether the present relearning data might be treated as Kott had, by adjusting the means for both learning and recall scores since these scores were themselves affected by experimental conditions. For this reason, although omitted from the body of this paper, the analysis of covariance data for relearning scores adjusted for both learning and recall scores have been deposited with the ADI (see Footnote 3) and are available for those who are interested. It may be noted that whereas the unadjusted groups × quality interaction was not significant, it became so at the .05 level when adjusted for learning scores, and significant at the .01 level when adjusted for both learning and recall scores.

do neurotics." The data, as measured by the relearning method, significantly support this hypothesis. "Neurotics retain more mildly toned than strongly toned material" was the third hypothesis and the findings support it significantly by both recall and relearning methods. Finally, the hypothesis that "normals retain more personally relevant strongly toned material than do neurotics" was similarly supported to a significant degree.

In the bottom half of the table, the data show that among normals, retention was significantly better for strongly toned than for mildly toned material. No other significant differences for any other comparisons were found except one: better retention by the relearning method of unpleasant material by the normals than pleasant material by the neurotics.

DISCUSSION

Effectiveness of Quality and Intensity Variables

More definitive data are available for intensity than for quality, in that there were more significant differences between the groups and stronger agreement between the two measures of retention. This observation is in line with Rapaport's summary statement after a review of recall experiments: "The intensity of the 'emotional factor' proved to be of more importance in its influence on memory than its quality . . ." (16, p. 94).

In addition to the general finding that intensity plays a more important role than quality, a pattern of effectiveness of intensity factors seems to emerge. Within the groups, neurotics retained mildly toned material more efficiently, whereas normals retained strongly toned material more efficiently. Between the groups, neurotics retained mildly toned material more efficiently than did the normals, and normals retained strongly toned material more efficiently than did the neurotics. In other words, at the time of remembering, the mildly toned material seemed to favor or be favored by the neurotics, while the strongly toned material seemed to favor or be favored by the normals. It may be assumed that all *S*s were affected by the emotionally toned material in that such material became more vivid as the intensity increased, resulting in better retention following the time lapse. However, when the material became too intense, a disruptive effect appeared, which counteracted any increasing vividness and resulted in loss of retention. Thus, it seems that two factors, vividness (making for efficiency of retention) and disruptiveness (making for inefficiency) were operating here. Furthermore, along the continuum of intensity, there were different points of maximum efficiency of retention, followed by disruptive reactions by both groups.

It appears that neurotics are more sensitive to the emotional intensity of a situation than are normals; mildly toned material becomes vivid for them sooner than it does for normals. As the emotional intensity gets stronger, disruptiveness begins to interfere with the increasing vividness that usually accompanies stronger intensity. For the normals, who appear to tolerate increasing intensity more easily than neurotics, vividness increases for a longer period as a function of intensity. However, even in their case a degree of intensity is reached at which the disruptive effect becomes operative.

In regard to quality, it was found that the pleasantly toned material failed to differentiate between the two groups, while the unpleasantly toned material resulted in relatively better retention by the normal group as measured by relearning. This trend may be attributable to a greater degree of emotional disturbance aroused in the neurotics during the time lapse and to the better ability of the normals to tolerate this material.

Relevance for Freud's Theory of Repression

Experimental studies of repression are subject to three types of evaluation. First, there is the claim that experimental studies can prove or disprove the theory (19). Second, there is the contention that the theory cannot be tested experimentally at all (13, 14). Between the two extremes is the point of view represented by Sears (16). Although pessimistic, he holds out some hope that laboratory investigations may contribute toward some "refinement of the theory," the possible "addition of relevant new variables," or the introduction of "techniques that promise eventual solution."

For those who believe that Freud's repression theory can be tested by means of experiments of this nature, the results of this study apparently fail to support the theory. In his

discussion of the theory of repression, Freud restricted his observations to qualitative factors, i.e. pleasantness-unpleasantness, with greater retention of the former over the latter. In the present study, as pointed out above, it is precisely with regard to the comparisons between pleasantly toned and unpleasantly toned material that little or no significant differences were obtained. Thus, for both groups, there was no support for the thesis that pleasantly toned material is better retained than unpleasantly toned material. Gordon, using a medium other than verbal material, obtained similar results and stated, "Pleasant odors are not more likely to be recalled than the unpleasant . . . . We must search for other factors than affective tone if we would understand why some impressions are recalled and others are not" (7, p. 225). According to the present results, retention appears to be based chiefly on the relationship between intensity factors and the emotional status of the *S*s.

In this experiment, as in Kott's (9), differential effects in retention became manifest, not at the time of original learning, but after a lapse of one week. Sharp (17) found similar results in that differences failed to occur until after the second day. Schmidt (15) likewise found that his more anxious *S*s did not differ in items correctly learned from his less anxious ones at the time of learning. One week later, differences in stimulus generalization responses occurred.

SUMMARY

The purpose of this investigation was to study the effects of affective intensity and quality of verbal material on the retention of this material by normal and neurotic male *S*s, testing hypotheses arising from Freud's theory of repression. The groups did not differ significantly as to age and schooling. Each *S* constructed personally meaningful two-word phrases, which he associated with nonsense syllables and learned to a criterion of two consecutive perfect repetitions within twenty trials. A week later he was required to recall and relearn them. Analysis of covariance was employed, with recall and relearning scores as the final measures and the original learning scores as the initial measures.

The results and their relation to the hypotheses may be summarized thus:

1. The data failed to support the hypothesis that "both normals and neurotics retain more pleasant than unpleasant material."

2. The data support the hypothesis that "normals retain more personally relevant unpleasant material than do neurotics" when retention is measured by the relearning method.

3. The data support the hypothesis that "neurotics retain more mildly toned than strongly toned material" by both recall and relearning methods.

4. The data support the hypothesis that "normals retain more personally relevant strongly toned material than do neurotics" by both measures of retention.

5. There were no significant differences for outcomes based on single variables. Thus, there were no significant differences in the retention of pleasant or unpleasant materials or of strong or mild material, and there were no differences with respect to single variables between the normal and neurotic groups.

6. The phrase-construction technique developed for this study appears to be a suitable means for obtaining personally meaningful verbal material for use in studying memory function as influenced by affective factors.

7. If Freud's theory of repression is at all testable by measuring retention of learned pleasant-unpleasant verbal material, then the obtained data do not support his hypothesis of motivated forgetting since neither group retained more personally relevant pleasant than unpleasant material. Instead, the present data suggest that forgetting is based on differential interactions between normal-neurotic grouping and affective intensity and quality factors, especially the former.

8. In investigations dealing with the relationship between personally meaningful material and memory function, it seems highly desirable to consider not only the variables of quality and intensity but also the variable of emotional stability in the *S*s.

REFERENCES

1. BARRETT, DOROTHY M. Memory in relation to hedonic tone. *Arch. Psychol., N. Y.*, 1938, No. 223. Pp. 61.
2. EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
3. FREUD, S. *Collected papers.* Vol. IV. London: Hogarth, 1950.

4. FREUD, S. *An outline of psychoanalysis*. New York: Norton, 1949.
5. GILBERT, G. M. The new status of experimental studies on the relationship of feeling to memory. *Psychol. Bull.*, 1938, **35**, 26–35.
6. GLAZE, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–269.
7. GORDON, KATE. The recollection of pleasant and unpleasant odors. *J. exp. Psychol.*, 1925, **8**, 225.
8. KLOPFER, B., & KELLEY, D. M. *The Rorschach technique*. New York: World Book Co., 1942.
9. KOTT, M. G. Learning and retention of words of sexual and nonsexual meaning. *J. abnorm. soc. Psychol.*, 1955, **50**, 378–382.
10. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. New York: Houghton Mifflin, 1953.
11. MELTZER, H. The present status of experimental studies on the relationship of feeling to memory. *Psychol. Rev.*, 1930, **37**, 124–139.
12. OSGOOD, C. E. Meaningful similarity and interference in learning. *J. exp. Psychol.*, 1946, **36**, 277–301.
13. RAPAPORT, D. *Emotions and memory*. Baltimore: Williams & Wilkins, 1942.
14. ROSENSTOCK, I. M. Perceptual aspects of repression. *J. abnorm. soc. Psychol.*, 1951, **46**, 304–315.
15. SCHMIDT, D. P. Some effects of manifest anxiety and symbolic reward and punishment on learning and stimulus generalization of complex visual material. Unpublished doctor's dissertation, Univer. of Washington, 1953.
16. SEARS, R. R. Survey of objective studies of psychoanalytic concepts. *Soc. Sci. Res. Coun. Bull.*, 1943, No. 51.
17. SHARP, AGNES A. An experimental test of Freud's doctrine of the relation of hedonic tone to memory revival. *J. exp. Psychol.*, 1938, **22**, 395–418.
18. THORNDIKE, E. L., & LORGE, I. *Teacher's word book of 30,000 words*. New York: Columbia Univer. Press, 1944.
19. ZELLER, A. F. An experimental analogue of repression. I. Historical summary. *Psychol. Bull.*, 1950, **47**, 39–51.

Received October 19, 1955.

REMINISCENCE, DRIVE, AND PERSONALITY THEORY[1]

H. J. EYSENCK

Institute of Psychiatry (Maudsley Hospital), University of London

SEVERAL attempts have been made in recent years to integrate personality study and learning theory. Of particular value in this connection has been the work of Mowrer (27), Dollard and Miller (4), and Spence (30, 31, 32). The reasons why these attempts cannot be regarded as wholly successful have been discussed elsewhere, and a "dynamic theory of anxiety and hysteria" (11) has been proposed, which attempts in a formal manner to relate the postulates of learning theory to an empirically founded and experimentally tested description of certain major personality variables.

This system of personality description has been reported in detail (7, 8, 9, 10). For the present purpose, it must suffice to say that two independent dimensions of personality (neuroticism and extraversion-introversion) have been identified quantitatively. These dimensions are related to classical psychiatry in the manner indicated by Jung: hysterics and psychopaths have high scores on both extraversion and neuroticism; whereas dysthymics, suffering from depression, anxiety states, or obsessional and compulsive symptoms, have high scores on both neuroticism and introversion. (It is not assumed, of course, that these two dimensions are the only ones which describe human conduct; what is maintained is simply that these are two very important dimensions, the understanding of which would make possible the description and control of a considerable proportion of human behavior.

PROBLEM

The postulate that links the dimensions of extraversion-introversion with modern learning theory and specifically that system of constructs used by Pavlov (29) and Hull (19) is called the typological postulate, and has been formulated thus:

> Individuals in whom reactive inhibition is generated quickly, in whom strong reactive inhibitions are generated, and in whom reactive inhibition is dissipated slowly, are thereby predisposed to develop extraverted patterns of behavior and to develop hysterico-psychopathic disorders in case of neurotic breakdown; conversely, individuals in whom reactive inhibition has developed slowly, in whom weak reactive inhibitions are generated, and in whom reactive inhibition is dissipated quickly, are thereby predisposed to develop introverted patterns of behavior and to develop dysthymic disorders in case of neurotic breakdown" (11, p. 35).

This postulate mediates a large number of predictions in the fields of learning theory and perception. One of these predictions, relating figural aftereffect to extraversion, has been verified in a previous paper (12, 28). Another prediction, relating conditioning to introversion, has been verified by Franks for both neurotic and normal *S*s (15, 16). The present paper is concerned with a third prediction, relating to the phenomenon of reminiscence. If, as is maintained with good evidence by Hull (19) and others (26), reminiscence phenomena are the product of inhibition originating in the massing of practice, then it follows from the typological postulate that *extraverts should show a higher degree of reminiscence than introverts.* The argument is probably too obvious to require much elaboration. Massed practice produces inhibition; inhibition is generated more quickly in extraverts; consequently, massed practice produces a greater amount of inhibition in the extraverted group. Reminiscence is produced by the dissipation of the accumulated inhibition during a rest period; reminiscence, therefore, is proportional to the amount of inhibition accumulated during massed practice. Extraverts who have accumulated more inhibition will therefore have more inhibition to dissipate and will consequently show greater reminiscence effects (11).

METHOD

Apparatus. The apparatus used was a traditional pursuit rotor. The turntable was 10 inches in diameter; the metal disc was $\frac{7}{10}$ inch in diameter, and the distance getween the center of the turntable and the center of the disc was $3\frac{1}{4}$ inches. The Turntable revolved clockwise at a speed of 72 r.p.s.; *S* attempted to keep the

[1] The author is indebted to the Research Fund of the Maudsley and Bethlem Royal Hospitals for generous help in making this investigation possible. He is also indebted to Mr. H. G. Jones for help on some of the theoretical points, and to Mr. Kolman Star, who carried out some of the computations.

tip of an articulated rod in contact with the metal disc, having been given the standard instructions to move the rod in a circular manner and not to press down on the turntable. The experiment was carried out in a blacked-out room with controlled artificial illumination equal for all *S*s. Recordings were made of the length of time during which the tip of the rod was in touch with the rotating disc during successive 10-second periods, an automatic device switching the recording between two recording chronometers every 10 seconds.

Procedure. There were three experimental sessions, divided from each other by 10-minute rest pauses. Each of the experimental sessions consisted of 30 successive 10-second trials, so that as far as the *S* was concerned he was working constantly for 5 minutes, had a break of 10 minutes, worked another 5 minutes, had another break of 10 minutes, and then worked another final 5 minutes. These conditions followed Ammons (2), who showed that this combination of massed practice and rest periods gives near optimum conditions for the emergence of reminiscence phenomena. Two breaks, rather than the more usual single break, were introduced in order to determine the reliability of the reminiscence score.

Subjects. The sample tested consisted of 50 experimentally naive male university students who were paid for their services. These students were volunteers who came to the Institute of Psychiatry for one day, during the course of which they were subjected to a number of different experiments. The pursuit rotor was applied relatively early during the day, so little generalized fatigue is likely to have been operative. No measures of intelligence were taken as the evidence tends to show that intelligence is not related to pursuit-rotor performance at IQ levels above 100. (The mean IQ level of similar groups used in previous experiments has usually been in the neighborhood of 120, with an *SD* of 10.) It is also known that correlations between intelligence on the one hand and neuroticism and extraversion on the other are very low or nonexistent. Little is known about the influence of age, but as the group was relatively homogeneous in this respect, few being below 19 or above 25, it does not seem likely that the results could have been influenced to any important extent by these variables.

Criterion tests. It is obviously crucial for the proper test of the hypothesis that a reasonably reliable and valid estimate of extraversion-introversion and neuroticism should be available. As has been argued elsewhere (9), the most desirable type of evidence comes from a battery of devices, including ratings, self-ratings, objective tests, and physiological measures. It was impossible on this occasion to apply a whole battery of this kind and a single measuring instrument was reluctantly used. This instrument was the Maudsley Personality Inventory.

This questionnaire consists of two scales containing 24 items defining extraversion-introversion and 24 items defining neuroticism, as well as buffer items and a lie scale. The extraversion-introversion scale has a corrected split-half reliability of .77; the neuroticism scale one of .88. The correlation between the two scales is −.05 (these figures are derived from the original standardization sample of 200 men and 200 women).

The scales were derived by means of an extensive item analysis from Guilford's STDCR and GAMIN questionnaires, and the Maudsley Medical Questionnaire. Some of the evidence relating to the development of the Maudsley Personality Inventory has been discussed elsewhere (9), but no detailed discussion of the final questionnaire has yet appeared.[2]

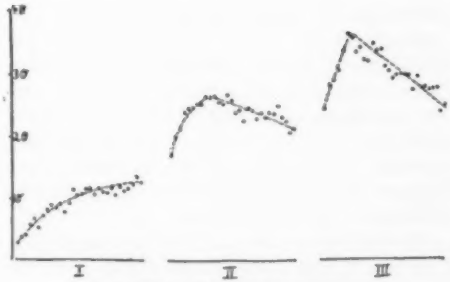

FIG. 1. MEAN PERFORMANCE ON THE PURSUIT ROTOR

The score plotted is percentage of time on target during successive 10-second trials in three 5-minute periods of massed practice, interrupted by two 10-minute rest pauses ($N = 50$).

RESULTS

Reminiscence. Figure 1 shows in graphic form the average scores of the 50 *S*s during the three sets of 30 trials. For ease of discussion, successive trials will be referred to by Roman numerals to indicate the first, second, and third sessions, and by Arabic numbers to indicate the trial within each experimental period. Thus, II-30 is the last (thirtieth) trial during the second period, and III-1 is the first trial during the third period. The figure shows that the main effects of experiments of this type are clearly present. These are (*a*) the reminiscence effect itself or the rise in performance from I-30 to II-1 and from II-30 to III-1; (*b*) a "warm-up" effect after the rest periods, continuing roughly from II-1 to II-10 and from III-1 to III-7; (*c*) a decline in performance after the peak of the "warm-up" period has been reached.

The score used here to represent individual differences in reminiscence was the difference between the average of the three last prerest trials and the first postrest trial. This measure does not take into account the practice effect of the last trial of a session, but this is probably more than compensated by not including the warm-up effect. The reason for taking the average of 3 scores as an estimate of prerest

[2] Copies may be obtained from the writer.

performance is simply the greater reliability of such an average as compared with a single 10-second trial score, coupled with the fact that performance appeared to have reached an asymptote. (This statement is more obviously true of I than of II.) Postrest performance cannot be measured in a similar manner because of the rapid rise of the curve (II-1 to II-10, and III-1 to III-7).[3]

Under these conditions, the correlation between the two reminiscence scores was .44 which, while highly significant, is clearly not very satisfactory. There are likely to be several reasons for this low reliability. The first and major one probably lies in the very high variance attaching to any single 10-second score. Chance factors, as is well known, play an enormous role in performance on the pursuit rotor. It is customary to get rid of such factors by averaging large numbers of trials or *S*s. To do so is not possible in the case of individual measurement, however, and as single 10-second trials play a considerable role in the formula used here for measuring reminiscence, there is obviously a considerable amount of error variance involved in this measurement.

Another difficulty which may account for the low correlation stems from the fact that during the performance of an experiment such as this, a considerable amount of ${}_{S}I_{R}$ accumulates. This does not dissipate during rest, although, as argued above, some of it is extinguished during the so-called warm-up period. The amount of conditioned inhibition which develops is a function of (*a*) the amount of reactive inhibition, the dissipation of which during involuntary rest pauses is supposed to act as a reinforcement; (*b*) the amount of drive, which determines the level reached by reactive inhibition before these rest pauses take place, and thereby indirectly the number of rest pauses during a given period; and (*c*) the conditionability of the *S*. Depending as it does on these three variables, conditioned inhibition must be considered to vary relatively independently of reactive inhibition, thus acting as a major source of error. Kimble (20, 21, 23, 24, 25) has shown that reactive inhibition develops early and tends to remain relatively constant, whereas conditioned inhibition develops late and increases in a negatively accelerated fashion. It follows that the second measure of reminiscence (RS_2) would be far more disturbed by the error variance introduced by conditioned inhibition than would RS_1. If this view is correct, it would explain in part the low correlation between the two, and it would also suggest that RS_1 should produce higher correlations with the indices of extraversion and neuroticism than RS_2.

Personality correlates. The correlation between extraversion and the first reminiscence score is .29, significant at the .02 level by an appropriate one-tailed test. The correlation with the second reminiscence score is .10, which is not significant. The expectation of higher correlations with RS_1 than with RS_2 is thus borne out in the absolute magnitudes, but the difference is not significant statistically. The correlation between extraversion and the sum of the two reminiscence scores is .22, which just fails of significance at the .05 level. The correlation between neuroticism and RS_1 is .40, significant at the .01 level. That between neuroticism and RS_2 is .27. The second score is lower than the first, but not significantly so. When the two reminiscence scores are summed, they correlate .34 with neuroticism. This correlation is significant at the .02 level. (The significance of the correlations between neuroticism and the reminiscence scores was assessed by means of a two-tailed test, as no prediction was made as to direction.)

The correlation between the questionnaire measures of extraversion and neuroticism in

[3] Ammons (1) has suggested a different method for the scoring of reminiscence, involving a correction for warm-up decrement. In an unpublished series of experiments, the writer has found evidence to suggest that Ammons's method of correcting for this decrement is based on a faulty theory of warm-up. The preferable alternative theory somewhat resembles that put forward by Denny *et al.* (3) to account for the fact that *S*s who changed from massed to distributed practice soon reached the performance level of *S*s who had been practicing under distributed conditions all the time. In brief, ${}_{S}I_{R}$ accumulates during the first period and, being a habit, is not dissipated during the first rest period. At the beginning of Period II, ${}_{S}I_{R}$ fails to be reinforced because of the absence of I_R at the beginning of the new testing period and after a long rest during which all reactive inhibition has been dissipated. Because of this failure of reinforcement, ${}_{S}I_{R}$ is extinguished, and consequently a steep rise in performance occurs. (The same explanation, of course, obtains in the case of the first seven trials in Period III.) If this account is accepted as a sound explanation, then clearly Ammons's method of measuring reminiscence confounds the effects of I_R and ${}_{S}I_{R}$ in proportions depending on the precise experimental arrangement.

this sample is quite insignificant (r = .115). It would therefore appear justifiable to conclude that about 25 per cent of the total variance of the reminiscence score RS_1 is predictable in terms of the two personality variables; only about 8 per cent of RS_2 is predictable in those terms. It would not be correct to say that personality factors, such as neuroticism and extraversion, *determine* the degree of reminiscence shown. Both reminiscence and personality factors are conceived of in this theory as determined by more fundamental causes, such as the inhibition/excitation balance, or the amount of autonomic drive present in a person. These are presumed to be largely innate features of the central and autonomic nervous systems respectively (13, 14) although this aspect of the theory is perhaps the one least strongly supported by direct experimental evidence. However, whether innate or acquired, these hypothetical properties of the nervous system are in this theory considered to underlie individual differences in *both* personality and learning.

Practice, decline, and warm-up. In view of the paucity of information on individual differences in various aspects of pursuit-rotor learning, it may be of interest to report some of the observed relationships. The first of these is the practice effect, defined arbitrarily as terminal score minus initial score, and obtained by subtracting the first three scores on the first 5-minute practice period from the last three scores. Theoretically, this effect would be facilitated by a strong excitatory potential and depressed by a strong inhibitory potential, so that it would be expected to correlate negatively with RS_1 and RS_2. The correlations are, in fact, both negative (−.48 and −.18). As, however, the sum of trials 28, 29, 30 in the first period enters into both RS_1 *and* the score of the practice effect, the higher of these two correlations may in part be an artifact, and as the lower one is not independently significant, no conclusions can be derived from these data.

The correlations of practice score with extraversion and neuroticism are both negative (−.103 and −.154); these are both in the expected direction, but cannot be interpreted as differing significantly from zero.

A decline score was calculated for Periods II and III by subtracting the sum of the three terminal scores (II-28, II-29, and II-30; and III-29, and III-30) from the sum of the highest III-28, three consecutive points reached in these two sessions. The hypothesis was that this decline would in part be due to reactive inhibition and should, therefore, be positively correlated with reminiscence. The two correlations in question were both positive (.06 and .31). The latter is almost significant at the .01 level, using a one-tailed test; the former is not significant. The reliability of this decline score is relatively low, with the correlation between the two decline scores only .13. It is likely that this low correlation is in part due to the differential influence of conditioned inhibition at different points on the learning curve.

A measure was taken of the warm-up effect by subtracting II-1 and III-1 respectively from the highest successive three scores obtained on Sessions II and III. These measures of warm-up show a just significant reliability, the correlation between the two being .32. Warm-up scores are negatively correlated with reminiscence, the warm-up in II correlating −.69 with RS_1 and −.35 with RS_2. The warm-up in III correlated −.30 with RS_1 and −.69 with RS_2. Again, the higher of these values are suspect as the same term enters into both sides of the correlation. Such is not the case, however, for the two smaller correlations, and it therefore seems to be established that the amount of warm-up observed is negatively related to the amount of reminiscence. The correlations between warm-up and both extraversion and neuroticism are all negative, but well below significance. It would be unsafe to assume that warm-up as measured here is related to either of these two dimensions of personality.

DISCUSSION

Little discussion appears to be necessary with respect to the relationship found between extraversion and reminiscence. The prediction made from the typological postulate was verified at an acceptable level of significance. It appears likely that a much closer relationship would be found if a more reliable and valid measure of extraversion were to be used and if a more reliable method of determining reminiscence could be devised. Experiments along these lines have indicated that shorter periods of practice (60 to 90 seconds) might give better results than the rather longer period employed in the present experiment.

Some discussion, however, appears to be necessary on the relationship between reminiscence and neuroticism. An explanation may be found along the following lines. First of all, it may be postulated with Mowrer (27), Dollard and Miller (4), Spence and Taylor (31, 32, 33), and other learning theorists that neuroticism is essentially an (autonomic) drive variable. The term usually chosen to identify this characteristic is, of course, *anxiety*. However, this term often implies not only what is psychiatrically known as anxiety state, but the whole group of neurotic disorders, so that "anxiety" and "neuroticism" are used interchangeably. This correspondence has been demonstrated by Franks (17), who found a correlation of .92 between the Maudsley Medical Questionnaire, a widely used measure of neuroticism (8), and the Taylor Anxiety Scale, a scale employed by Spence and his associates as an operational definition of anxiety. Hysterics are supposed to be psychiatrically differentiated from anxiety states in that they show relatively little anxiety. Yet hysterical *S*s had scores on the Taylor Anxiety Scale intermediate between those of a group of anxiety patients and a group of normals (17). Consequently, it may be clarifying to talk about neuroticism as the major variable characterizing all neurotic groups, and to use the term anxiety to characterize the introverted neurotic group. There is ample evidence to justify this usage (9).

The relevance of drive (in this case autonomic drive, or neuroticism) to reminiscence follows from an extension of the Hullian theory made by Kimble (22). According to him, the amount of reactive inhibition which *S* will tolerate in a motor-learning situation is in part a function of his drive; "the greater amount of this motivation the greater amount of I_R he will tolerate" (p. 248). Kimble explains that this hypothesis follows from the fact that reactive inhibition is thought of as a negative drive or as a need to rest.

> As such it acts in opposition to the needs leading to the continuation of practice. If we adopt the position that a given amount of I_R cancels a certain amount of these more positive needs, then it follows that the highly motivated *S* will continue to practice in spite of the presence of an amount of inhibition which would produce a cessation of activity in a less highly activated *S*. *The experimental prediction to which this argument leads is that the amount of reminiscence . . . will increase with an increase in motivation* (p. 248).

Both Kimble himself (22) and Wasserman (34) have provided experimental evidence in favor of this hypothesis, which may, therefore, be regarded as essentially substantiated.

The linkage of this hypothesis within learning theory to personality study is as follows. If one regards neuroticism as a drive, and if drive leads to an increase in the amount of I_R tolerated by the organism before producing a cessation of activity, then individuals with strong drives should produce greater amounts of I_R and consequently higher reminiscence scores. As this is precisely what is empirically found, one may regard this explanation as quite feasible. It would be interesting to know whether neurotics as compared with normals have higher reminiscence scores. A prediction to this effect would seem to follow from Kimble's theory and the results here reported. However, the situation may be more complicated because the drive in this connection would presumably have to be a relevant one, and it would be rather more difficult to ensure an equal relevance of autonomic drive as between hospitalized neurotics and nonhospitalized normals than it would be between *S*s having high and low neuroticism scores respectively as in this study.

Among neurotics, one is probably on safe ground in predicting that hysterics and psychopaths will have higher reminiscence scores than anxiety states and reactive depressives (dysthymics). In terms of the MMPI and the use made of the *Hy-Pt* score devised by Eriksen (5, 6) one would probably also be correct in predicting that this score would correlate positively with reminiscence in both neurotics and normals. Thus, putting the various groups in order of expected reminiscence-score size, one has hysterics and psychopaths, followed by normal extraverts and dysthymics, with normal introverts last. It is impossible to predict with any confidence the relative positions of normal extraverts and of dysthymics; such a prediction would require a more precise knowledge of the exact degree of relationship obtaining between I_R and neuroticism and extraversion, respectively.

SUMMARY

This paper reports an experiment performed to test the hypothesis derived from learning theory, that *reminiscence on the pursuit rotor*

following massed practice should be greater for extraverts than for introverts. The relationship between reminiscence and neuroticism was also investigated. Fifty university students were employed as subjects, using pursuit-rotor learning as a measure of reminiscence and the Maudsley Personality Inventory as a measure of extraversion-introversion and neuroticism. Extraversion was found to be associated with reminiscence, according to hypothesis, and reminiscence was also found to be significantly greater for subjects high in neuroticism than for subjects low in neuroticism. These results are interpreted as indicating the fruitfulness of linking personality study with learning theory (11).

REFERENCES

1. AMMONS, R. B. Acquisition of motor skills; I. Quantitative analysis and theoretical formulation. *Psychol. Rev.*, 1947, **54**, 263–281.
2. AMMONS, R. B. Acquisition of motor skills: II. Rotary pursuit performance with continuous practice before and after a single rest. *J. exp. Psychol.*, 1947, **37**, 393–411.
3. DENNY, M. R., FRISBY, N., & WEAVER, J. Rotary pursuit performance under alternate conditions of distributed and massed practice. *J. exp. Psychol.*, 1955, **49**, 48–54.
4. DOLLARD, J., & MILLER, N. E., *Personality and psychotherapy.* New York: McGraw-Hill, 1950.
5. ERIKSEN, C. W. Psychological defences and "ego-strength" in the recall of completed and incompleted tasks. *J. abnorm. soc. Psychol.*, 1954, **49**, 45–50.
6. ERIKSEN, C. W. Some personality correlates of stimulus generalization under stress. *J. abnorm. soc. Psychol.*, 1954, **49**, 562–566.
7. EYSENCK, H. J. *Dimensions of personality.* London: Kegan Paul, 1947.
8. EYSENCK, H. J. *The scientific study of personality.* London: Routledge and Kegan Paul, 1952.
9. EYSENCK, H. J. *The structure of human personality.* London: Methuen, 1953.
10. EYSENCK, H. J. *The psychology of politics.* London: Routledge & Kegan Paul, 1954.
11. EYSENCK, H. J. A dynamic theory of anxiety and hysteria. *J. ment. Sci.* 1955, **101**, 28–51.
12. EYSENCK, H. J. Cortical inhibition, figural after-effect, and the theory of personality. *J. abnorm. soc. Psychol.*, 1955, **51**, 94–106.
13. EYSENCK, H. J. The inheritance of extraversion-introversion, *Acta Psychologica*, 1956, **XII**, 95–110.
14. EYSENCK, H. J., & PRESSL, D. B. The inheritance of neuroticism: an experimental study. *J. ment. Sci.*, 1951, **97**, 441–465.
15. FRANKS, C. M. Conditioning and personality: A study of normal and neurotic subjects. *J. abnorm. soc. Psychol.*, 1955, **2**, 143–156.
16. FRANKS, C. M. Personality factors and the rate of conditioning. *Brit. J. Psychol.*, 1957, in press.
17. FRANKS, C. M. An experimental study of conditioning as related to mental abnormality. Unpublished doctor's dissertation, Univer. of London, 1954.
18. HILDEBRAND, H. P. A factorial study of introversion-extraversion by means of objective tests. Unpublished doctor's dissertation, Univer. of London, 1953.
19. HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
20. KIMBLE, G. A. An experimental test of a two-factor theory of inhibition. *J. exp. Psychol.*, 1949, **39**, 15–23.
21. KIMBLE, G. A. Performance and reminiscence in motor learning as a function of the degree of distribution of practice. *J. exp. Psychol.*, 1949, **39**, 500–510.
22. KIMBLE, G. A. Evidence for the role of motivation in determining the amount of reminiscence in pursuit rotor learning. *J. exp. Psychol.*, 1950, **40**, 248–253.
23. KIMBLE, G. A., & BILODEAU, G. A. Work and rest as variables in cyclical motor learning. *J. exp. Psychol.*, 1949, **39**, 150–157.
24. KIMPLE, G. A., & HORENSTEIN, B. R. Reminiscence in motor learning as a function of length of interpolated rest. *J. exp. Psychol.*, 1948, **38**, 239–244.
25. KIMBLE, G. A., & SHATEL, R. B. The relationship between two kinds of inhibition and the amount of practice. *J. exp. Psychol.*, 1952, **44**, 355–359.
26. MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning.* (2nd Rev. Ed.) London: Longmans Green, 1952.
27. MOWRER, O. H. *Learning theory and personality dynamics.* New York: Ronald, 1950.
28. NICHOLS, E. G. The relation between certain personality variables and the figural after-effect. Unpublished doctor's dissertation, Univer. of London, 1956.
29. PAVLOV, I. P. *Conditioned reflexes.* London: Oxford Univer. Press, 1927.
30. SPENCE, F. W., & FARBER, I. E. Conditioning and extinction as a function of anxiety.
31. SPENCE, K. W., & TAYLOR, JANET A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.
32. SPENCE, K. W., & TAYLOR, JANET A. The relation of conditioned response strength to anxiety in normal, neurotic, and psychotic subjects. *J. exp. Psychol.*, 1952, **45**, 265–272.
33. TAYLOR, JANET A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.
34. WASSERMAN, H. N. The effect of motivation and amount of pre-rest practice upon inhibitory potential in motor learning. *J. exp. Psychol.*, 1951, **42**, 162–172.

Received December 19, 1955.

PERCEPTUAL ORGANIZATION AND INTELLIGENCE: A FURTHER STUDY[1]

CLINTON DE SOTO[2] AND H. LEIBOWITZ

University of Wisconsin

SINCE the time of Galton (1), attempts have been made to measure intelligence with sensory discrimination tasks. These efforts have generally shown little success, and intelligence testing has consequently followed a different course. A new approach to the study of intelligence has been suggested by Krech and Calvin (2), who report a positive correlation between the ease with which *S*s proceed through levels of perceptual organization and intelligence as measured by the Wechsler vocabulary test. (6). The purpose of the present study is to extend their experiments and to determine the effect of stimulus duration on this perceptual performance.

Krech and Calvin performed two experiments, with 28 college students as *S*s in the first and 19 in the second. The apparatus for the perceptual task consisted of a Harvard tachistoscope, stimulus cards, a "reproduction board," and small black beads. The stimuli were patterns composed of black dots ($\frac{1}{16}$ inch squares) on a white background. The first experiment employed two patterns, an X figure and a rectangular array. The second experiment employed two rectangular arrays. The rectangular arrays included 72 dots linearly arranged in a 6×12 pattern (the entire pattern was approximately square). The dots were actually arranged in 6 rows of 12 dots per row, with $1\frac{1}{64}$ inch between neighboring rows and $\frac{3}{64}$ inch between dots within rows. In the experiment using two rectangular arrays the rows ran vertically for one pattern ("columns") and horizontally for the other. The patterns were separated by $\frac{1}{2}$ inch. The reproduction board contained 900 concave hemispheres ($\frac{1}{2}$-inch diameter) arranged closely in a 30×30 pattern.[3]

The *S*s were told the test was of visual acuity and they were shown how to reproduce the stimulus by placing beads in the depressions on the board. They were told that they would repeatedly be presented briefly with two symbols, one of which was correct. They were to guess which was correct until they solved the problem and to reproduce both symbols after each presentation. The problem-solving demand was expected to elicit in each *S* the highest perceptual organization of which he was capable.

The stimuli were presented for 0.06-second exposures. The position order of the two patterns varied randomly over trials. The solution to the problem was a LRLR . . . alternation sequence. Trials continued until *S* achieved 6 consecutive correct choices or until 24 trials (16 trials in the second experiment) were completed.

Their results show that 28 *S*s spaced the beads so as to reproduce rows and columns for some or all trials, while 19 *S*s persistently reproduced the patterns as homogeneous patterns of equidistant dots. According to Krech and Calvin's theoretical discussion, the former (we will call them "high perceivers") showed a higher level of perceptual organization than the "low perceivers." In both experiments, high perceivers tended to achieve higher vocabulary scores than low perceivers (a mean of 33.6 as compared with 27.9 in the first experiment, and 30.6 as compared with 27.4 in the second experiment; these differences were significant as shown by the *t* test and a nonparametric test). A vocabulary score of 30.5 almost perfectly divided the high from the low perceivers. Further evidence that row and column reproduction is a high achievement is that few *S*s, once attaining it, reverted to homogeneous reproductions.

Their theoretical discussion suggests that had the distance between the row groupings been greater or had the exposure time been longer, more of the *S*s would have achieved the row grouping concept. In other words, they regard it as a threshold phenomenon.

[1] This research was supported with funds provided by the Wisconsin State Legislature and by research grant M-1090 from the National Institute of Mental Health of the National Institutes of Health, Public Health Service.

[2] Now at Johns Hopkins University.

[3] Dr. Allen D. Calvin supplied us with a sample stimulus from which these specifications were estimated.

The authors performed an experiment similar to Krech and Calvin's, varying stimulus exposure time and using somewhat modified experimental techniques.

METHOD

Thirty-six students from introductory psychology classes at the University of Wisconsin served as *Ss*. All had normal near vision as determined by a reading chart test.

A Harvard tachistoscope[4] and stimulus patterns were employed which were essentially the same as in Krech and Calvin's second experiment with the exception that the two patterns of dots were separated by 1¼ inches rather than ½ inch. A single center dot was presented in the adapting field of the tachistoscope and served as a fixation point prior to each stimulus exposure. The *Ss* observed binocularly at a viewing distance of 23 inches.

The *Ss* drew the stimulus by making dots with a number 2 pencil in the squares of graph paper uniformly ruled with light brown lines, 20 to an inch. This procedure was adopted to allow more accurate representations of the stimulus, particularly in size, than the reproduction board. It was assumed that "ability to draw" is scarcely more involved in placing dots in squares than in placing beads in holes.

We omitted the instruction to *S* to learn to choose the correct symbol and emphasized solely the importance of reproducing accurately what he had seen (calling it a "visual acuity" experiment, as did Krech and Calvin). The left-right position of the row and column patterns varied randomly over trials. The *Ss* drew both patterns on each trial.

Three exposure times were used: 0.01, 0.10, and 1.00 second. Illuminance level was constant at 10 foot-candles. Twelve *Ss* were assigned randomly to each exposure-time condition. Each *S* was given 16 trials at one of the three exposure times, then a 17th trial on which the stimulus was exposed for 30 seconds.

The Wechsler vocabulary test was administered prior to the perception test and the *Ss* were told that the two tests belonged to different research programs. The vocabulary scoring was done later for all *Ss* anonymously and simultaneously.

RESULTS

The over-all distribution of vocabulary score (mean = 29.17, s = 3.23) was similar to the distributions obtained by Krech and Calvin (mean = 31.98, s = 4.29; mean = 28.71, s = 2.70). The present sample is thus suitable in this respect for further study of their findings. The mean vocabulary score varied nonsignificantly among the exposure-time conditions.

The present results are similar to those of Krech and Calvin's in that the same proportion of *Ss*, 60 per cent, were high perceivers (reproduced rows and columns accurately on at least one trial). We verified their finding that *Ss* tend to progress from homogeneous reproductions to row-column ones and rarely revert.

TABLE 1
PERCEPTUAL PERFORMANCE AND VOCABULARY SCORE

Stimulus Exposure Time	High Perceivers	Low Perceivers	diff.	t
0.01 sec.				
N	9	3		
Mean Wechsler Vocabulary Score	28.56	26.17	2.39	1.09
0.10 sec.				
N	6	6		
Mean Wechsler Vocabulary Score	29.75	30.50	−0.75	.37
1.00 sec.				
N	7	5		
Mean Wechsler Vocabulary Score	29.00	30.00	−1.00	.57

However, we found no relationship between perception and vocabulary scores, nor did we find a relationship between perception scores and stimulus-exposure time. Table 1 shows the number (N) of high and low perceivers in each exposure-time condition, together with their mean vocabulary score, the difference between the means of the high and low perceivers in each condition, and the ts for these differences. For each condition, a t of 2.23 would be required for significance at the .05 level, consequently none of the obtained ts were significant. We examined the operating characteristic function[5] for t for the .05 confidence level, based upon 10 df, using a separate estimate (s) for each condition (s = 3.29, 3.50, and 2.98 for the 0.01-, 0.10-, and 1.00-second conditions, respectively.) The O-C function indicates that if the true difference between the mean vocabulary scores of the high and low perceivers had been 3.76 in the

[4] Manufactured by Ralph Gerbrands, Arlington, Massachusetts.

[5] Selection of the .05 significance level results in a 5 per cent chance of Type I errors (rejection of the null hypothesis H_0 when H_0 is true). In addition, use of the t-test, with 10 df, gives a certain chance of Type II errors (acceptance of H_0 when H_0 is false, i.e., failure to detect a true difference between the population means). The probability of Type II errors in case various absolute differences between means exist is given by the operating-characteristic (O-C) functions of the t test, using the sample estimates of σ. The O-C functions of a test are complements of its power functions. For some discussion of this see (5).

0.01-second condition, the probability of accepting H_0 (failing to reject it) would be .10. In other words, 90 per cent of repeated experiments would show a significant difference. If the true difference had been 3.99 in the 0.10-second condition, the probability of accepting H_0 would be .10. If the true difference had been 3.39 at 1.00 second, the probability of accepting H_0 would be .10. If the true difference had been 5.20 at 0.01 second, 5.53 at 0.10 second, and 4.70 at 1.00 second, the probability of accepting H_0 would be only .01 for each condition.

Because duration of exposure did not appear to affect the relative proportion of high and low perceivers (see the χ^2 test below), another t test was done between the high and low perceivers' mean vocabulary scores with the exposure groups combined. The high perceivers' mean was 29.02, and low perceivers' 29.39. The t for this difference is .33, far from significance. The O-C function for t for the .05% confidence level, $df = 34$, $s = 3.28$, shows that if the true difference were 1.87, the probability of accepting H_0 would be .10; and if the true difference were 2.53, the probability would be .01.

On Trial 17, with a 30-second exposure time, 28 *Ss* reproduced rows and columns and 8 did not. The mean vocabulary score of the high perceivers was 29.44; of the low perceivers, 29.09. This difference gives a t of .26, far from significance. The O-C function here (.05 level, $df = 34$, $s = 3.28$) is the same as that just described for all *Ss* on Trials 1–16 because s was almost identical. Again, if the true difference were 1.87, the probability of accepting H_0 would be .10; and if it were 2.53, the probability would be .01.

All these conjectural true differences, the first three fully independent, are smaller than the difference between means of 5.72 which Krech and Calvin found in their first experiment. They compare well with the difference of 3.15 obtained in their second experiment. Since the vocabulary distributions and the percentage of high perceivers were similar to theirs, it appears that each of our experimental conditions should have provided a fairly powerful test of the perception-intelligence hypothesis. However, our results were uniformly negative.

A possible relationship between time of exposure and the percentage of high perceivers (or the intelligence cutoff value between high and low perceivers) was suggested by Krech and Calvin. Table 1 (the N entries) shows the number of high and low perceivers in each exposure time condition. The χ^2 for this table is 1.636, with $.30 < p < .50$, far from significance. Exposure time over the range from 0.01 to 1.00 second does not appear to affect perception of rows and columns. Such a finding is consistent with the previous observation that stimulus localization in the frontal plane is independent of the duration and luminance of the stimulus or stimuli (3, 4).

Only one of the 22 high perceivers on Trials 1–16 reverted to low perceiving on Trial 17, while 7 of the 14 low perceivers on Trials 1–16 became high perceivers on Trial 17. However, the correlated proportions χ^2 for these frequencies, corrected for continuity, is 3.13, not significant ($.05 < p < .10$).

The data were examined for sex differences (both males and females were used in all conditions) but none appeared.

American Council Psychological Examination (ACE) percentile ranks were obtained for most of the *Ss*. Kendall's rank-correlation coefficient Tau is .49 for the ACE total scores correlated with the vocabulary scores and .54 for the ACE L scores correlated with vocabulary. The ACE scores showed no relationship to perception scores.

Casual observations during the experiment suggest that many *Ss* who failed to reproduce rows and columns on the graph paper might have been credited with seeing them according to other measures. Some *Ss*, despite the instructions, attempted to draw dashes or solid lines following the stimulus rows and columns. When told to draw dots only, these *Ss* sometimes produced homogeneous dot patterns and received no credit for rows and columns. Other *Ss* spontaneously reported seeing rows and columns which they could not draw, attributing the rows and columns to factors other than spacing, such as the angle of light incidence and the angle of view. Many *Ss* were observed to draw the dots in horizontal rows (with lateral hand movements) for the row pattern and vertically for the column pattern, yet produce two identical homogeneous dot patterns. A few *Ss*, however, did report they had seen no difference or no internal difference between the patterns for some or all the trials.

DISCUSSION

The failure to repeat Krech and Calvin's findings demands careful study of the experimental techniques involved. The *S* population and method of stimulus presentation appear to be essentially identical. The instructions were somewhat different, those of the present study emphasizing accurate reproduction of the stimulus rather than learning the correct symbol. The percept-measuring responses were definitely different, although both would seem to involve the same type and level of critical abilities. Krech and Calvin's reproduction board allowed a maximum of 2 or 4 dots to an inch, while our graph paper allowed 10 or 20 to an inch. (The stimulus showed 9 dots to an inch within rows.) Perhaps this difference in the available sizes of the stimulus reproduction is important.

In any case, it would appear that the crucial factor or factors do not operate simply to make high perceptual organization easier or more difficult, but rather randomize it with respect to intelligence, and perhaps with respect to exposure time.

The facts that in the present experiment many low-vocabulary *S*s did reproduce rows and columns, that 9 of 12 *S*s did so even at the short exposure time of 0.01 second, and the indications that many *S*s who did not reproduce them on graph paper were nevertheless perceiving them, suggest that almost all if not all *S*s actually perceive them. If this is true, then any consistent failure of low-intelligence *S*s to reproduce rows and columns may be a phenomenon more characteristic of the response than of the perceptual mechanism. The failure of 8 *S*s to reproduce rows and columns even after a 30-second stimulus exposure bears out a response interpretation.

An alternative perceptual hypothesis may be suggested by the *S*s who evidenced row-column percepts by reports and drawing techniques some trials before they actually reproduced rows and columns. Particularly striking are the 7 *S*s who first reproduced them following the 30-second exposure even though many gave earlier evidence of seeing them. These *S*s apparently could perceive rows and columns on brief exposure but could discover the spacing characteristics responsible only on a long exposure. Perhaps the perception of spacing determinants is a third level of perceptual organization, following the homogeneous dots percept and a simple row-column percept.

Some elaboration of explanation such as those proposed seems required by the puzzling effects on measured perception of intelligence, stimulus duration, and response measures. The correlation of perception with intelligence (2), certainly a remarkable finding, appears to be revealed by the present study as complexly or obscurely determined and difficult to interpret.

SUMMARY

Krech and Calvin (2) obtained a significant positive correlation between intelligence and level of perceptual organization, which was measured by the ability to reproduce the rows and columns of patterns of dots viewed for 0.06 second. The present experiment, using modified techniques, showed no such correlation between intelligence and perception, although available evidence indicates the statistical tests possessed adequate power to test the hypothesis of such a correlation.

Krech and Calvin found that progression over trials was usually from low to high perceptual organization, rarely the reverse. Our data confirmed this finding. Contrary to hypothesis, the data show no correlation between perception and stimulus-exposure time from 0.01 to 1.00 second, and perhaps even through 30 seconds. There were indications in our data that many *S*s who failed to demonstrate high perceptual organization with our stimulus-reproduction technique might have done so with other techniques. Possible interpretations of the data are discussed.

REFERENCES

1. GALTON, F. *Inquiries into human faculty and its development.* London: Macmillan, 1883.
2. KRECH, D., & CALVIN, A. Levels of perceptual organization and cognition. *J. abnorm. soc. Psychol.*, 1953, **48**, 394–400.
3. LEIBOWITZ, H., MYERS, N. A., & GRANT, D. A. Radial localization of a single stimulus as a function of luminance and duration of exposure. *J. opt. Soc. Amer.*, 1955, **45**, 76–78.
4. LEIBOWITZ, H. W., MYERS, N. A., & GRANT, D. A. Frequency of seeing and radial localization of single and multiple visual stimuli. *J. exp. Psychol.*, 1955, **50**, 369–373.
5. MOSTELLER, F., & BUSH, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology.* Cambridge: Addison-Wesley, 1954. Pp. 289–334.
6. WECHSLER, D. *The measurement of adult intelligence.* (2nd Ed.) Baltimore: Williams & Wilkins, 1941.

Received January 9, 1956.

IMPRESSIONS OF PERSONALITY, AUTHORITARIANISM, AND THE *FAIT-ACCOMPLI* EFFECT[1]

I. H. PAUL

Austen Riggs Center, Stockbridge, Massachusetts

A UNITED STATES presidential election is an event of social impact and importance; a time when prominent political figures vie publicly for a position of conspicuous leadership, and people are sensitive to these figures as persons. This study is an investigation of people's impressions during the 1952 U. S. presidential election of the personalities of these leader-figures: how these impressions are affected both by the outcome of the election and by the people's orientation to authority.

Persons often face situations that, initially conflictful and equivocal, have become structured and settled by an accomplished fact beyond their control. Social psychologists have studied this situation during events such as elections and major legislative actions, and they repeatedly found that the *fait accompli* influences most people's opinions and attitudes and results in increased acceptance of the new state of affairs (3, 11). How general are the cognitive changes? Are impressions of the personalities of contending candidates in an election also affected by a *fait accompli*? The present study shows that they are.

Although the perception of another's personality is a crucial pivot in interpersonal dealings, it has only recently begun to attract the interest of psychologists. Pepitone has shown how a person who is seen to possess power tends, under certain conditions, to be viewed as approving, benevolent, and generally attractive (14). The valence of a person—his perceived attractiveness—seems also to be readily affected by external events (9), as well as by status and group membership (8), and it is well known that an important action of a person or one associated with him can radically change our impression of him.

Winning a U.S. presidential election is an important event: it results in a conspicuous change in status for the winner and endows him with power and success. Therefore, the first major expectation in this study is that the election outcome is followed by a change in people's impressions of the presidential candidates' personalities and results in an increase in the winner's attractiveness.

What role do a person's needs and dispositions play in such changes of impressions? Asch (2) recognized that the "needs and claims" which relate the perceiver to the perceived person may play a fundamental part in impressions of personality.[2] Such needs and claims, in the case of a person's relation to authority figures, are embodied in the concept of the "authoritarian personality." The conception—a broad and heuristically valuable one—has already stimulated extensive research.

Jones (10) recently reported an experiment that demonstrates the role authoritarianism plays in the formation of the person's first impressions of a leader-figure. Authoritarian persons tended to evaluate the leader more favorably than equalitarians did and, while in general they were less sensitive to the psychological or personality characteristics of others, they seemed more sensitive to such external variables as social status or position, and especially to power (defined as institutionally sanctioned authority).

The authoritarian person tends to conform and submit to authority (1, 12, 17). One of his submitting tactics might presumably consist of changing his impression of the person to whom he needs to submit. Such a tactic, in psychoanalytic terms, may be part of a rationalization or, in Festinger's terms (6), an attempt to bring his cognitions into *consonance* with a changed situation. A central assumption is that the type of change in impression which promotes such submission or consonance is, in

[1] This article is based on a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Pennsylvania. The writer wishes to thank Professor M. G. Preston who gave generous and invaluable supervision to every phase of this study. He also thanks Professor J. Wishner and A. Pepitone, members of his dissertation committee, for their guidance; and Miss M. Molitch for her assistance in collecting data.

[2] A study by Preston, *et al.* (15) which shows that married persons' impressions of their spouses' personalities are a function of the marriage relationship itself is relevant to this point.

part, an increase in the favorability or attractiveness of that impression.

Therefore, the second major expectation in this study is that the authoritarian person in particular comes to view the winning candidate as an attractive person, i.e., he is more susceptible to the *fait-accompli* effect than is his equalitarian neighbor. A further expectation is that the authoritarian's alleged tendency to dichotomous thinking should result in greater differences between the ratings of his favored and his nonfavored leader-figures.

METHOD

Selection of groups. Three weeks before the election day in November of 1952 a short questionnaire-battery was administered to a pool of 320 volunteer business-school students of an introductory psychology course at the University of Pennsylvania. This battery consisted of: (*a*) The Short Authoritarian-Equalitarian Scale;[3] (*b*) a statement of candidate-preference; (*c*) an index of preference-strength (*"Indicate the strength of your feeling"*); (*d*) an index of ego-involvement (*"If your candidate does not win, how badly will you personally feel?"*); (*e*) an open-ended question about the reasons for candidate-preference. A coding technique was used to maintain subject anonymity while facilitating the selection.

The results of this pretesting showed that 170 *S*s favored Eisenhower and 150 Stevenson. There proved to be no relationship between A-E score and candidate-preference.

Four groups were selected from the upper and lower quartiles of the A-E distributions such that the two authoritarian groups—the one endorsing Eisenhower, the other Stevenson—had comparable A–E scores; and the same was true of the two equalitarian groups. Of 114 students asked to participate in the experiment, 96 appeared for the pre-election testing and, of these, 87 returned for the postelection testing. Seven *S*s were later eliminated in order to equalize the four experimental groups at 20 *S*s each.

The results of the pretesting showed that the groups differed in their preference-strength and ego-involvement. The proportion of authoritarian pro-Eisenhower *S*s who professed *very strong* feelings was greater than the corresponding proportion of authoritarian pro-Stevenson *S*s ($CR = 2.11$, $p < .05$). The same was true for ego involvement. The differences between the equalitarian groups, though not as great, were consistent and in the opposite direction. These findings, of importance in their own right, are discussed below. However, the results on these items for the combined authoritarian groups did not differ from those for the combined equalitarian groups. The authoritarian *S*s were somewhat more ego involved but this difference is not statistically significant.

Pre-election testing. In pre-election testing four to five days preceding the election day, the following battery was administered: (*a*) Impressions-of-Personality scales pertaining to Stevenson, Eisenhower, and *S* himself; (*b*) a repeat of the Short A-E Scale, and of the indices of candidate preference, preference strength, and ego involvement; (*c*) a number of biographical items such as age, socioeconomic status, and political orientation.[4]

Since the time of initial testing, no *S* had changed his candidate preference or his A-E classification, and the retest results for the items dealing with preference strength and ego involvement remained virtually identical.

Postelection testing. The postelection testing in which the Impressions-of-Personality scales were readministered took place two to three weeks following the election day. At the time of postelection testing, Eisenhower had only stepped into his new role but was not yet functioning in it, though he had made known two of his prospective cabinet choices—Dulles and Wilson. Stevenson ostensibly took his defeat with grace and good feelings.

Impressions-of-Personality scales. The work of Cattell (4) and of Fiske (7) on the factorial structures of personality ratings supplied a basis for devising the scale of impressions of personality. Twelve items were selected from Fiske's scale, 12 from Cattell's, and four new ones were added. The trait-items were classified into five a priori groups or classes, the first four being based upon the conclusions of Cattell and of Fiske: I Social Adaptability, II Emotional Control, III Conformity, IV Inquiring Intellect, and V Leadership Strength.[5]

For each candidate, *S* indicated his impression on each trait by placing a check mark in the appropriate position along an 8-point scale whose extremes are labeled and defined. To allow this rating to be judged with respect to favorability or attractiveness, *S* indicated, in a space beside each trait-item, at which pole he felt *"a President ought to fall,"* i.e., which is the

[3] The Short A-E Scale, devised and standardized by Sanford and Older (18), is composed of items adapted from the California F scale plus other items, and was designed to meet the needs of a study of persons' general orientation to leadership. The scale was found to correlate .67 with the California F scale and the authors report a repeat reliability of .78 and split-half reliability of .58. In a subsequent study (17), a close relationship was demonstrated between a person's score on the scale and his overt orientation to leaders and to leadership phenomena. Recently, Eager and Smith (5) have reported data obtained in a children's summer-camp situation, which speak strongly for the predictive power of this scale.

[4] The self-impressions and biographical items will not be considered in the present report.

[5] Examples of each trait class: I, *Attentive to People* ("Interested in people, their troubles, their personalities. Makes friends with people and remembers their personal interests"); II, *Unshakable Poise* ("Self-possessed. Does not lose composure under emotional provocation"); III, *Independent-Minded* ("Thinks things out for himself. Adopts a clear and definite independent position"); IV, *Broad Interests* ("Has wide interests. Well informed on a wide variety of subjects. Shows intellectual curiosity"); V, *Persevering* ("Persists in any enterprise undertaken in spite of opposition. Steadfast").

TABLE 1

MEAN PRE- AND POSTELECTION RATINGS OF CANDIDATES BY AUTHORITARIANISM AND CANDIDATE PREFERENCE (20 CASES IN EACH SUBGROUP)

Trait-Class	Time	Authoritarian			Equalitarian			Total
		Pro-Eis.	Pro-Ste.	Total	Pro-Eis.	Pro-Ste.	Total	
		Ratings of Eisenhower						
I. Social Adaptability	Pre-election	5.73	5.41	5.58	6.05	5.79	5.92	5.74
	Postelection	6.34	6.12	6.23	6.24	6.08	6.16	6.20
II. Emotional Control	Pre-election	6.08	5.10	5.59	6.40	5.88	6.14	5.87
	Postelection	6.56	6.07	6.32	6.66	6.37	6.52	6.42
III. Conformity	Pre-election	6.35	4.92	5.63	6.19	5.36	5.78	5.70
	Postelection	6.52	5.60	6.06	6.45	5.67	6.06	6.06
IV. Inquiring Intellect	Pre-election	6.22	4.50	5.38	5.95	5.03	5.49	5.44
	Postelection	6.53	5.58	6.06	6.26	5.70	5.98	6.02
V. Leadership Strength	Pre-election	6.21	5.32	5.77	6.09	5.53	5.81	5.79
	Postelection	6.47	6.10	6.29	6.34	5.91	6.14	6.21
Combined Trait-scale	Pre-election	6.12	5.05	5.58	6.13	5.52	5.82	5.70
	Postelection	6.49	5.90	6.19	6.39	5.94	6.17	6.18
		Ratings of Stevenson						
I. Social Adaptability	Pre-election	4.98	6.06	5.53	5.65	6.29	5.97	5.75
	Postelection	5.42	6.36	5.89	5.83	6.20	6.02	5.95
II. Emotional Control	Pre-election	5.21	6.73	5.97	6.43	6.75	6.59	6.28
	Postelection	5.83	6.84	6.34	6.71	6.80	6.75	6.54
III. Conformity	Pre-election	5.13	6.41	5.77	5.58	6.75	6.17	5.97
	Postelection	5.45	6.33	5.89	6.04	6.58	6.31	6.10
IV. Inquiring Intellect	Pre-election	5.11	6.38	5.75	5.75	6.63	6.19	5.97
	Postelection	5.59	6.61	6.10	6.32	7.11	6.71	6.45
V. Leadership Strength	Pre-election	5.27	6.40	5.84	5.99	6.65	6.33	6.08
	Postelection	5.61	6.31	5.95	6.18	6.40	6.28	6.12
Combined Trait-scale	Pre-election	5.14	6.39	5.77	5.88	6.61	6.25	6.01
	Postelection	5.58	6.49	6.03	6.24	6.61	6.41	6.22

"desirable" rating. Each rating was then scored so that the 8 end represented the high-valued pole. The instructions, which were both printed and spoken, urged the *S*s to reveal their impressions of the personality of the leader-figures rather than merely to give political opinions.

RESULTS

The ratings by the members of each experimental group were pooled to yield mean ratings for each candidate on each of the five classes of trait-items, as well as on the trait-scale as a whole. Table 1 presents these mean ratings in the pre-election and postelection tests.

The data were treated by analysis of variance.[6] Since none of the quadruple-interactions proved significant when tested against the quintuple-interaction, they were combined with it to furnish a composite highest-order-interaction variance.[7] This variance then served as error-term to the triple-interactions. In testing each double-interaction and primary source the error-term selected was the next higher order interaction involving *S* or trait-classes (if that interaction itself proved significant). Table 2 presents the results of the analysis of variance.

1. Findings with respect to the expectation

[6] The experimental design is a factorially confounded one: whereas each *S* rates both of the candidates in both testings, the *S*s who compose the four experimental groups—the endorsement and authoritarian-equalitarian groups—are different. Therefore, higher-order-interaction variances involving combinations of *S*s, as a variable, with these two variables cannot be extracted. Since the design furnishes no replication data, careful attention needs to be paid to the proper selection of error-terms (higher-order-interactions) for the most valid assessment of statistical significance. It should be noted that the confounding eliminates some of the most sensitive error-terms that involve *S*s as a variable, and interactions involving trait-classes have to be used in their place.

[7] For a discussion of this procedure, see McNemar (13, p. 315). It should be noted that the quadruple-interaction involving *S* cannot be combined with the others, nor can it validly be tested against the quintuple-interaction which does not involve *S*. Therefore it remains as the highest-order-interaction to be used in assessing those sources in which *S* figures.

TABLE 2
ANALYSIS OF VARIANCE OF THE MEAN RATINGS

Source	*df*	Variance Estimate	Error Term	*F* Ratio	*p*
1. Of Eisenhower and Stevenson (L-F)	1	120.58	*S* 11	1.75	
2. Pre- and Postelection (I-F)	1	474.38	*S* 15	42.43	.001
3. Pro-Eisenhower & Pro-Stevenson (End)	1	22.13	*S* 34	2.77	
4. Authoritarian & Equalitarian (A-E)	1	290.25	*S* 34	36.46	.001
5. Trait-Classes (Tr)	4	69.05	*S* 19	9.93	.001
6. Subjects (*S*)	76	66.34	*S* 19	9.54	.001
7. L-F × I-F	1	63.63	*S* 23	7.70	.01
8. L-F × End	1	2258.26	*S* 25	54.32	.01
9. L-F × A-E	1	104.17	*S* 34	13.09	.01
10. L-F × Tr	4	36.52	*S* 27	6.43	.001
11. L-F × *S*	76	68.93	*S* 27	12.13	.001
12. I-F × End	1	1.09	*S* 29	3.11	
13. I-F × A-E	1	31.59	*S* 34	3.97	<.05
14. I-F × Tr	4	10.91	*S* 27	2.79	.05
15. I-F × *S*	76	11.18	*S* 31	2.86	.01
16. End × A-E	1	12.49	*S* 34	1.56	
17. End × Tr	4	10.07	*S* 34	1.26	
18. A-E × Tr	4	8.79	*S* 34	1.11	
19. Tr × *S*	304	6.95	*S* 31	1.78	.01
20. L-F × I-F × End	1	109.39	*S* 34	13.74	.01
21. L-F × I-F × A-E	1	16.60	*S* 34	2.09	<.10
22. L-F × I-F × Tr	4	1.39	*S* 33	3.52	.01
23. L-F × I-F × *S*	76	8.27	*S* 33	20.67	.001
24. L-F × End × A-E	1	113.01	*S* 34	14.19	.01
25. L-F × End × Tr	4	41.57	*S* 34	5.22	.01
26. L-F × A-E × Tr	4	4.58	*S* 34	1.73	
27. L-F × Tr × *S*	304	5.68	*S* 33	14.27	.001
28. I-F × End × A-E	1	4.66	*S* 34	1.70	
29. I-F × End × Tr	4	3.69	*S* 34	2.16	
30. I-F × A-E × Tr	4	4.51	*S* 34	1.76	
31. I-F × Tr × *S*	304	3.91	*S* 33	9.33	.001
32. End × A-E × Tr	4	8.65	*S* 34	1.09	
33. L-F × I-F × Tr × *S*	304	0.40			
34. Combined quadruple- and quintuple-interactions	21	7.96			

that people's impressions of the personalities of presidential candidates change following the election and this change involves an increase in the attractiveness of the winner:

Table 1 shows that for all *S*s the increase in ratings on the combined trait-scale was greater in their impressions of Eisenhower than of Stevenson. While it is true that Stevenson, having been rated higher in the pre-election testing, continued afterwards to be rated very slightly higher (the difference is not statistically significant), the increment from the pre- to the postelection testing is significantly greater in the mean rating of Eisenhower.[8]

[8] When dealing with changes along a linear scale, attention needs to be paid to the possible presence and influence of a scale-effect, because the poles of such a scale may exert a limiting influence on such change. In the case of the present study, since increase in rating is the preponderant finding and since the ratings are characteristically in the upper half of the scale, each finding needed to be scrutinized carefully to make sure it could not be due to a "ceiling-effect." It may be noted that the analysis of variance provides, in the interaction of I-F × *S*s, a partial index to scale-effect which can be used as the error-term where such an effect is suspected. Several other attacks on the problem were also made. Each of the major findings reported in this study has withstood rigorous scrutiny with regard to a "ceiling-effect."

The ratings of Eisenhower go from 5.70 to 6.18, an increase of .48, while those of Stevenson go from 6.01 to 6.22. an increase of .21 (the interaction L-F × I-F, *S* 6 in Table 2, represents this finding).

Inspection of Table 1 shows that in each of the five trait-classes the increase, for the experimental population as a whole, was greater in the impressions of Eisenhower than of Stevenson, though in the case of Inquiring Intellect (IV) it was virtually the same for both. In all the cases the pre-election rating of Eisenhower was the lower, but in two cases (Social Adaptivity and Strong Leadership) the postelection ratings of Eisenhower rose beyond those of Stevenson.

The differential increase in favor of Eisenhower varied with the trait-classes (L-F × I-F × Tr). It was greatest in the trait-class Strong Leadership (.38). The rank order of the remaining four is: Emotional Control (.29), Social Adaptability (.26), Conformity (.23), and Inquiring Intellect (.08).

The influence of the *fait accompli* was greater on those *S*s who endorsed the loser (L-F × I-F × End). Table 1 shows that, while the ratings of Eisenhower increased more than those of Stevenson in the case of each group's rating regardless of the candidate initially favored, the relationship was stronger in the case of the unfavored candidate. In the latter case the ratings of Eisenhower on the combined trait-scale were initially lower but became slightly higher than those of Stevenson. Further examination of Table 1 reveals that ratings became less favorable in two instances (Conformity and Strong Leadership), both in ratings of Stevenson by those *S*s who endorsed him. Moreover, this group showed a greater increase in its rating of Stevenson on Inquiring Intellect. The greatest increase in rating occurred for Eisenhower by pro-Stevenson *S*s (.64); the next greatest for Stevenson by pro-Eisenhower *S*s (.40); the

next for Eisenhower by pro-Eisenhower *S*s (.32); and the least for Stevenson by pro-Stevenson *S*s (.04, which is not significantly different from zero: $t = 0.29$).

2. Findings with respect to the expectation that the differential increase in favor of the winner is greater among authoritarian than among equalitarian persons.

It has been seen that ratings of Eisenhower changed in a favorable direction to a greater extent than those of Stevenson. Such differential change in favor of the winner was more pronounced among authoritarian than among equalitarian *S*s. The difference between these groups (L-F × I-F × A-E), however, does not attain statistical significance beyond the .10 level of confidence. Nevertheless, in the case of the authoritarians but not of the equalitarians, Eisenhower, initially lower, emerged finally higher than Stevenson. The differential increase in favor of Eisenhower was .35 among the authoritarians, and only .19 among the equalitarians.

Table 1 shows that in four of the five trait-classes the increment in the ratings of Eisenhower was greater than in that of Stevenson, both for the authoritarian and the equalitarian groups. The exception is the trait-class Inquiring Intellect where the increment of Stevenson was the greater for the equalitarian groups. A reversal of position (Eisenhower beginning as the lower of the two but emerging in the postelection as the higher) occurred only for one trait-class (Social Adaptability) among the equalitarian *S*s, while it appeared in three trait-classes in the ratings of the authoritarian *S*s (Social Adaptability, Conformity, and Strong Leadership). Change in an unfavorable direction appeared in only one instance, the rating of Stevenson by equalitarian *S*s on Strong Leadership. While the rating of Stevenson on this trait-class by the authoritarian *S*s increased very slightly, the difference is not statistically significant.

In the ratings of their nonfavored candidate the greater favorable change in impressions of Eisenhower was most pronounced for the authoritarian *S*s. Among these *S*s appeared the most conspicuous example of the *fait-accompli* effect. Moreover, their initial ratings of Eisenhower were lower than of Stevenson and their final ratings higher in all trait-classes except one (Inquiring Intellect, where the final ratings are practically the same).

3. Further findings. Ratings of favored candidates were higher than ratings of nonfavored candidates (L-F × End). Moreover, the discrepancies between *S*s' ratings of favored and nonfavored leader-figures were greater in the pre-election testing than in the postelection testing (L-F × I-F × End). In both periods these discrepancies were greater for authoritarian *S*s than for equalitarian *S*s[9]—the mean pre-election discrepancy for authoritarian *S*s being 1.16, for equalitarian *S*s, 0.67; in the postelection testing, 0.75 and 0.45, respectively. Inspection of Table 1 suggests that this finding, which holds consistently for each of the trait-classes, is primarily due to the authoritarians' lower ratings of their nonfavored leader-figure. In each instance the authoritarian *S*s rated their nonfavored candidate less favorably than did their equalitarian fellow-endorsers.

DISCUSSION

The results of this study bear out both of the major expectations. The *fait-accompli* effect is reflected in the finding, demonstrated to some extent in each of the groups, that the ratings of Eisenhower's personality changed in the direction of favorability or attractiveness more than did the ratings of Stevenson. Furthermore, the extent of this differential change was greater among the authoritarian *S*s than among the equalitarians.

The *fait-accompli* effect was also greatest on those *S*s, both authoritarian and equalitarian, who originally endorsed the loser. It stands to reason that this group should be most affected by the election outcome since each of its members was at variance with majority opinion and choice and had acquired a president whom he had viewed as a relatively unattractive person. The fact that his impression of the winner showed more improvement may be understood as an attempt to reach an equilibrium between his inclinations and the reality created by the election outcome—a "facilitative perception," a measure to attain "cognitive consonance," or a "rationalization."

Most of the *S*s expressed awareness of the

[9] However, the interaction which represents this finding (L-F × A-E × End) does not attain statistical significance. This may largely be due to the absence of a sufficiently sensitive higher-order-interaction variance which could serve as its error-term.

change in their impressions and many of them attributed it directly to the election outcome and its implications. Here are a few representative examples of their responses to the open-ended question concerning felt change:

> I feel a little more confident in his ability as president. The change . . . is probably due only to the reason that he has been elected president. I can't explain myself too well, but I feel that Eisenhower is just a better man than I had imagined.
>
> Personal feelings of Eisenhower have changed. Nothing that could be easily and categorically written down. I feel he is more of a man because of his victory.
>
> The change in my impressions slightly toward Eisenhower is probably based on the respect to be shown to his office and the dire hope that he will turn out to be a better president than I thought he would have made.
>
> Eisenhower being our hope for the next four years, becomes (out of desperation?) a better man.
>
> I think I respect Ike more now, probably because: 1) he won by such a majority; 2) he is the president-elect; 3) he is picking fairly decent men for the top government posts.

The relationship between endorsement and favorableness of impression is clear-cut: positive endorsement goes with a more favorable view. However, in this connection some additional findings deserve consideration: namely, that in the post-election testing impressions of *both* leader-figures became more attractive, and those of *S*'s initially nonendorsed candidate (whether winner or loser) did so to a greater extent.

Thomsen (19) found a shift in attitude intensity following the 1938 election from a U-shaped distribution toward a normal curve, which he interpreted in terms of a general decrease in intensity of attitude following the election, with a shift toward a more neutral position. It seems reasonable to assume, therefore, that the intensity of ego involvement with regard to candidate endorsement should also appreciably subside after the election. If we further assume that, in the tense and exciting atmosphere of the pre-election period, heightened ego involvement and attitude strength are influential primarily in bringing about a devaluation of a person's opposed candidate, then it follows that when this atmosphere is alleviated equilibrium may be re-established by improvement of this impression. This formulation, in addition to explaining the over-all improvement in the ratings of non-endorsed candidates, also leads to the important prediction that the impressions of the loser by those *S*s who initially favor him, should in fact not increase. We expect an increase for the winner by those who endorse him because of the various perceptual and motivational changes which are associated with the *fait-accompli* effect: he is a figure crowned with success and power. But in the case of the loser there is no reason to expect that those who initially endorse him should emerge with a more favorable impression. The data bear out this expectation. They show that it is precisely these latter impressions which showed no significant change.[10]

The role of ego involvement and attitude strength are further brought into prominence by the following findings. The selection procedure revealed that, although authoritarians and equalitarians were distributed equally in both endorsement groups, the authoritarian pro-Stevensonites felt less strongly and were less ego involved in their endorsement of Stevenson than were their equalitarian fellow-endorsers. Conversely, of the pro-Eisenhowerites it was the authoritarians who felt more strongly and were more ego involved. Now, in their impressions of the candidates it is the authoritarians, and especially those who initially endorsed Stevenson, who changed most in favor of Eisenhower. It appears, therefore, that strength of endorsement and extent of ego involvement are factors which were involved, at least partly, in the role which authoritarianism played in the *fait-accompli* effect. That these factors cannot alone account for the main findings is suggested by a number of considerations, among which is the observation that there is no consistent relationship in the results between the ratings by *S*s and the extent of their ego involvement and strength of attitude.[11]

The proposition concerning the authoritarians' tendency to think dichotomously has support in the finding that there was a greater discrepancy between their ratings of the candidates than between those of the equalitarians. Rokeach's (16) contention that ego

[10] That these ratings fail to decrease may also need explanation. One may appeal, as Korchin does (11), to cultural factors in the U.S. which lend a loser some support and favor. Such values and mores are usually summed up in expressions like "on the side of the underdog," "sportsmanlike losing," and the like.

[11] For example, in the pre-election testing the authoritarian pro-Stevenson *S*s rated Stevenson no differently than the authoritarian pro-Eisenhower *S*s rated Eisenhower (see Table 1).

involvement is an important condition for the occurrence of this phenomenon is supported in the finding that this difference was greater in the pre-election period—a period which is most likely characterized by heightened ego involvement. These findings may be related to the authoritarians' devaluation of their nonfavored candidate in the pre-election period since, following the election outcome, the decrease in disparity seems to have resulted primarily from an increase in these ratings.

SUMMARY

The results of this study permit the conclusion that, in a leadership contest such as a U.S. presidential election, a *fait-accompli* effect occurs with respect to impressions of the personalities of the candidates. Furthermore, this effect is more pronounced among authoritarian persons than among equalitarians.

Ratings of the personalities of Eisenhower and Stevenson were obtained from authoritarian and equalitarian *S*s one week before the 1952 election and again several weeks after the election. When these ratings are judged with respect to attractiveness or favorableness, the findings show a greater postelection change in a favorable direction for Eisenhower. This differential increase in favor of the winner was found for *S*s who initially endorsed Eisenhower, for those who endorsed Stevenson, as well as for authoritarians and equalitarians. However, differential favorable change was greatest among the authoritarian *S*s and among those who had endorsed the loser. The greatest differential increase occurred regularly for the trait-class Leadership Strength.

A further finding is that the discrepancy between a person's ratings of his favored and nonfavored candidates was greater before the election than after, and greater for the authoritarian person than for the equalitarian.

REFERENCES

1. ADORNO, W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. ASCH, S. E. Forming impressions of personality. *J. abnorm. soc. Psychol.*, 1946, **41**, 258–290.
3. CANTRIL, H. *Gauging public opinion*. Princeton: Princeton Univer. Press, 1944.
4. CATTELL, R. B. The description of personality: basic traits resolved into clusters. *J. abnorm. soc. Psychol.*, 1943, **38**, 476–506.
5. EAGER, JOAN, & SMITH, M. B. A note on the validity of Sanford's authoritarian-equalitarian scale. *J. abnorm. soc. Psychol.*, 1952, **47**, 265–267.
6. FESTINGER, L. The relation between behavior and cognition. Unpublished paper read at Colo. Sympos. on Cognition, Boulder, 1955.
7. FISKE, D. W. Consistency of the factorial structures of personality ratings from different sources. *J. abnorm. soc. Psychol.*, 1949, **44**, 329–345.
8. HAIRE, M., & GRUNES, WILLA. F. Perceptual defenses: processes protecting an organized perception of another's personality. *Hum. Relat.*, 1950, **3**, 403–412.
9. HOROWITZ, M. W., LYONS, J., & PERLMUTTER, H. V. Induction of forces in discussion groups. *Hum. Relat.*, 1951, **4**, 57–75.
10. JONES, E. E. Authoritarianism as a determinant of first-impression formation. *J. Pers.*, 1954, **23**, 107–127.
11. KORCHIN, S. J. Restructuration of attitude following a national election: the *fait-accompli* effect. *Amer. Psychologist*, 1948, **3**, 272. (Abstract)
12. MASLOW, A. H. The authoritarian personality. *J. soc. Psychol.*, 1943, **30**, 29–48.
13. MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1949.
14. PEPITONE, A. Motivational effects in social perception. *Hum. Relat.*, 1950, **3**, 57–77.
15. PRESTON, M. G., MUDD, E. H., PELTZ, W. L., & FROSCHER, H. B. Impressions of personality as a function of marital conflict. *J. abnorm. soc. Psychol.*, 1952, **47**, 326–337.
16. ROKEACH, M. Generalized mental rigidity as a factor in ethnocentrism. *J. abnorm. soc. Psychol.*, 1948, **43**, 259–278.
17. SANFORD, F. H. *Authoritarianism and leadership: a study of the follower's orientation to authority*. Philadelphia: Institute for Research in Human Relations, 1950.
18. SANFORD, F. H., & OLDER, H. *The short authoritarian-equalitarian scale*. Philadelphia: Institute for Research in Human Relations, 1950.
19. THOMSEN, A. What voters think of candidates before and after elections. *Publ. Opin. Quart.*, 1938, **2**, 269–274.

Received January 10, 1956.

RELATIONS OF TEMPERAMENT TO THE CHOICE OF VALUES

LYLE V. JONES AND CHARLES MORRIS

The University of Chicago

FIVE dimensions in the domain of value were isolated in a previous study, and empirical data from several nationality groups were interpreted in terms of that dimensional structure (2). The data were obtained from administration of the Morris *Ways to Live* schedule to male college students in India, Japan, China, Norway, and the United States. The *Ways to Live* schedule requires that a subject assign a rating to each of 13 paragraphs, each defining a "way to live." Each way to live represents a conception of desirable life, such as those found in prominent religious or ethical traditions. The document has been presented in its entirety by Morris and Jones (2).

In contrast with the earlier study, the present investigation focuses upon individual differences in expressed values among members of the same cultural group. In particular, it is hypothesized that the expressed value system of an individual is related to his temperament. For the purpose of the investigation, the value system is determined from response to the Ways to Live document, and scores for the areas appraised by the *Thurstone Temperament Schedule* (5) are employed as indices of temperament traits.

An important distinction should be emphasized between the value and the temperament domains, as here measured. The *Ways to Live* document assesses an individual's belief concerning the way he would like to live. In contrast, items of the temperament schedule refer to aspects of actual behavior. There is no necessary relationship between the two sets of responses. It remains an empirical issue to determine the pattern of relationships which may appear between temperament and values.

METHOD

Description of Variables

The Thirteen Ways to Live. Each subject is instructed to assign a numeral from one to seven to each of the paragraphs designating a way to live in accordance with the degree to which he would like to live that kind of life. These assigned numerals serve as scores for the subject.[1] In Table 1 appear phrases serving to summarize the content of each of the *Ways to Live* paragraphs.

The Five Value Factors. Approximately 650 male college students of ages 20 or 21, from 21 colleges and universities throughout the United States, responded to the *Ways to Live* document during the period 1945 to 1954. Represented are both state and larger private universities. Within a school, students were chosen predominantly from general undergraduate courses in philosophy, the social sciences, and physical education, courses containing students with major interests in all parts of the university. The final sample of 250 was selected to be representative of this group.[2]

Product-moment correlation coefficients were computed among all pairs of ways as rated by these 250 persons; they appear in Table 2. A centroid factor analysis was performed and factors were rotated graphically to orthogonal simple structure (4, 6). In Table 3 is the centroid matrix, and the resulting rotated factor matrix is given in Table 4.[3]

The five value factors were then interpreted in terms of the factor solution of Table 4. A summary serving to identify the interpreted factors appears in Table 5. The thirteen variables in Table 1, together with the five variables of Table 5, define the value measures used in the present study.

The Seven Temperament Areas. In Table 6 are summarized the seven areas appraised by the *Thurstone Temperament Schedule.* From responses to the temperament schedule, a score may be determined for each subject on each of the seven temperament variables. The score is found by adding (*a*) the number of items in each temperament area for which response (yes or no) corresponds to the scoring key, and (*b*) one-half the number of items in that temperament area for which response is neither yes nor no.

Relating Values and Temperament

Both the *Ways to Live* document and the *Thurstone Temperament Schedule* were administered in 1951, 1953,

[1] For the purpose of obtaining product-moment correlations with other variables, it has been demonstrated that results from scores so defined do not differ substantially from results based upon the successive intervals scale values corresponding to the assigned numerals (cf. 2).

[2] Strictly speaking, the 250 were chosen from the larger group by a cyclic rather than a random sampling process. However, it is judged to be unlikely that bias was introduced.

[3] Since this is a different sample from that analyzed earlier (2), which neither was selected from students throughout the United States nor was controlled for age, the factor loadings here reported differ somewhat from those presented in the earlier paper.

and 1954, to an additional 115 male undergraduate students from general classes at Harvard University, University of Chicago, Illinois Institute of Technology, and Illinois Wesleyan University. As a preliminary survey of relationships between value and temperament, product-moment correlation coefficients were calculated between the seven temperament variables and the thirteen ways to live.

In order to ascertain the relationship between the temperament areas and the value dimensions, a procedure (1) was employed for estimating the factor loadings of variables not included in the original analysis. The method allows for estimation of the loadings of the value factors upon each of the seven temperament variables. In geometric terms, these loadings represent the projections of a vector associated with each temperament area upon the unit reference vectors representing the five orthogonal value factors.

TABLE 1
SUMMARY OF THE 13 WAYS TO LIVE

Way	Summary of Content
1	Refinement, moderation, restraint; preservation of the best attainments of man.
2	Self-sufficiency, understanding of self; avoidance of outward activity.
3	Sympathy, concern for others; restraint of one's self-assertiveness.
4	Abandonment, sensuous enjoyment of life; solitude and sociality both are necessary.
5	Energetic, cooperative action for the purpose of group achievement and enjoyment.
6	Activity; constant striving for improved techniques to control nature and society.
7	Flexibility, diversity within self; accept something from all other paths of life.
8	Carefree, relaxed, secure enjoyment.
9	Quiet receptivity to nature yields a rich self.
10	Dignity, self-control; but no retreat from the world.
11	Give up the world and develop the inner self.
12	Outward, energetic activity; use of the body's energy.
13	Let oneself be used; remain close to persons and to nature.

RESULTS AND DISCUSSION

In Table 7 appear the correlations between ways of life and temperament areas. It is apparent that several columns (ways to live) and several rows (temperament areas) exhibit no significant correlation coefficients. Thus, for the values associated with ways to live numbers 1, 3, 4, and 10, there is no evidence of relationship with temperament. And for temperament areas A and E, there are observed no significant relationships with value.

The significant correlations of Table 7 may be organized so as to illustrate certain trends of relationship. Ways 5, 6, 8, and 12 exhibit similar patterns of correlations with the temperament variables. They tend to be related positively to temperament areas I, V, S, and D, and negatively to R. A second distinct cluster of ways to live consists of Ways 2, 7, 9, and 11, which tend to correlate negatively with temperament areas I, V, S, and D, and positively with R. A negative relationship between Way 13 and temperament variable R exhausts the significant correlations of Table 7.

The method of Dwyer (1, 3) was employed to estimate the loadings of the five value factors upon the temperament variables. It is instructive to consider the resulting matrix of estimated factor loadings (Table 8) as an addition of seven rows to the factor matrix of Table 4, and to review the definitions of the value factors in light of these added results.

Value Factor A, Social Restraint and Self

TABLE 2
PRODUCT-MOMENT CORRELATION COEFFICIENTS AMONG THE 13 WAYS TO LIVE
(N = 250)

Test	1	2	3	4	5	6	7	8	9	10	11	12
1												
2	044											
3	197	−061										
4	−216	122	−109									
5	−008	−266	097	047								
6	010	−085	017	−100	133							
7	−014	099	−083	075	−119	004						
8	012	−010	013	274	118	−075	077					
9	−104	221	145	−034	−055	−029	056	084				
10	265	059	130	−224	−010	067	−176	−097	−002			
11	−057	278	−024	118	−255	−189	098	−009	153	130		
12	029	−153	033	080	176	237	−074	001	−011	107	−078	
13	047	−034	177	−203	029	003	−154	−099	276	156	112	057

Note.—Decimal points are omitted in all entries.

TABLE 3
CENTROID FACTOR MATRIX

	I	II	III	IV	V	h^2
1	291	215	−078	132	326	265
2	−410	359	078	259	157	397
3	250	208	136	−155	162	177
4	−416	−327	286	045	−055	374
5	312	−326	207	−186	066	290
6	291	−185	075	147	098	159
7	−293	−051	−083	059	124	111
8	−213	−207	223	−113	259	216
9	−148	290	400	−228	112	332
10	341	363	101	244	012	313
11	−343	370	158	183	−162	336
12	309	−233	356	254	−067	346
13	237	394	270	−245	−149	363
$\frac{\Sigma a^2}{13}$	093	083	047	035	025	(283)

Note.—Decimal points are omitted in all entries.

TABLE 4
ROTATED FACTOR MATRIX

	A	B	C	D	E	h^2
1	51	−03	−05	02	−02	264
2	10	−28	50	−11	20	392
3	25	01	−05	34	−03	181
4	−44	12	10	−11	38	374
5	−04	34	−36	20	07	292
6	17	30	−18	−06	02	156
7	−07	−19	02	−18	20	113
8	−12	02	−08	09	44	223
9	−03	−09	26	47	19	334
10	41	14	22	09	−25	306
11	−10	−17	54	−02	−02	331
12	04	58	00	−01	01	339
13	07	02	15	51	−27	361
$\frac{\Sigma a^2}{13}$	059	056	066	055	046	(282)

Note.—Decimal points are omitted in all entries.

Control, contributes little to the temperament variables. The factor accounts for less than two per cent of the variance of the temperament scores ($\Sigma\ a^2/7 = .019$). The largest loading, −.27, appears on temperament variable V. A negative relationship here is reasonable, suggesting that the physically vigorous individual is unlikely to ascribe great virtue to social restraint or self control.

Factor B, represented by Ways 12, 5, and 6, defined as Enjoyment and Progress in Action, exhibits high estimated loadings on temperament variables Impulsive, Vigorous, Sociable, and Dominant. These results tend to reinforce the summary interpretation of Factor B (Table 3), and suggest that those individuals who profess to value a purposive expenditure of energy also are those who do live vigorously, who are quick to make decisions, who are sociable, and who tend to exhibit initiative and qualities of leadership.

Factor C, Withdrawal and Self-Sufficiency, had been defined by Ways 2, 11, 9, and −5. This factor now appears to be positively related to Reflective temperament, and exhibits smaller negative loadings on Sociable, Vigorous, and Impulsive temperament areas. Again, the behavioral correlates, from the temperament domain, seem consistent with the related expressions of value.

Value Factor D, Receptivity and Sympathetic Concern, exhibits only one moderately strong loading on the temperament variables, a negative loading on Reflective. It is reasonable that the reflective individual, who prefers theoretical thinking and who works alone, regards unfavorably a life devoted to the service of others.

Value Factor E, which was given the name of Self-Indulgence, exhibits near-zero loadings on all temperament areas, with the possible exception of Reflective (.21). That the factor contributes to the Reflective area probably could be attributed to the affinity of the reflective individual for Ways 2, 4, 7, and −13 (Table 7). A similar pattern is seen among the Factor E loadings upon these Ways (Table 4).

The communality estimates of Table 8 suggest that the Thurstone Temperament variables I, V, S, R, and D share a substantial portion of variance with the value domain, but that temperament variables E (Stable) and A (Active) are relatively independent of

TABLE 5
SUMMARY OF VALUE-FACTOR INTERPRETATIONS

Factor	Interpretation
A	Ways 1, 10, −4, 3 Social restraint and self-control
B	Ways 12, 5, 6, −2 Enjoyment and progress in action
C	Ways 11, 2, −5, 9 Withdrawal and self-sufficiency
D	Ways 13, 9, 3 Receptivity and sympathetic concern
E	Ways 8, 4, −13, −10 Self-indulgence

TABLE 6
SUMMARY OF THE SEVEN THURSTONE TEMPERAMENT VARIABLES

Variable	Summary of Content
A: Active	Hurried in speech, walk, writing, driving and working; tends to be "on the go"; restless.
V: Vigorous	Physically active in sports and outdoor occupations; uses large muscle groups; tends toward great expenditure of energy.
I: Impulsive	Quick to make decisions; enjoys competition, easily changes from one task to another; tends to be daredevil, happy-go-lucky, carefree.
D: Dominant	Excels at public speaking, organizing social activities, persuading others; capable of taking initiative and responsibility; tends to be a leader.
E: Stable	Exhibits cheerful, even disposition; able to remain relaxed and calm in a crisis; does not fret about daily chores.
S: Sociable	Sympathetic, cooperative, receptive, and agreeable with others; makes friends easily and enjoys company of others.
R: Reflective	Prefers theoretical to practical problems; enjoys work requiring accuracy and detail; tends to be quiet, to work alone.

TABLE 7
PRODUCT-MOMENT CORRELATION COEFFICIENTS BETWEEN THURSTONE TEMPERAMENT VARIABLES AND WAYS TO LIVE
(N = 115)

Temperament Variable	Way 1	2	3	4	5	6	7	8	9	10	11	12	13	
A	095	−102	−055	−108	152	114	−145	031	001	147	−073	096	084	Active
V	−005	**−252**‡	−138	033	**275**	*187*	001	*196*	**−275**	007	**−319**	**480**	004	Vigorous
I	−034	*−210**	−048	003	**375**	**271**	**−279**	**282**	**−251**	076	**−247**	**395**	−019	Impulsive
D	145	*−191*	−008	068	**251**	164	*−224*	089	*−236*	161	−124	*187*	−011	Dominant
E	−024	−032	−124	−124	003	010	037	093	−075	041	015	068	037	Stable
S	005	**−336**	050	065	**400**	172	*−227*	*210*	*−200*	109	*−210*	*228*	105	Sociable
R	−007	**302**	−118	126	**−273**	−027	*184*	−104	076	−024	**246**	−120	**−299**	Reflective

Note.—Decimal points are omitted in all entries.
* *184*, Significant beyond .05
‡ **340**, Significant beyond .01

TABLE 8
ESTIMATED VALUE FACTOR LOADINGS ON THURSTONE TEMPERAMENT VARIABLES

Temperament Variable	Value Factor A	B	C	D	E	h^2
A	15	24	−06	09	−07	097
V	−27	69	−26	−13	−06	637
I	07	75	−24	−04	13	644
D	17	48	−15	−08	00	288
E	01	06	00	−06	−06	011
S	−02	55	−33	17	00	441
R	06	−20	38	−41	21	400
$\frac{\Sigma_s^2}{7}$	019	239	058	033	010	(359)

Note.—Decimal points are omitted in all entries.

value, as measured by the *Ways to Live* document.

The final row of Table 8 indicates the estimated proportion of variance of the temperament variables which could be accounted for by each of the value factors. Factor B contributes strongly to the temperament domain. Factors C and D contribute less strongly, while Factors A and E are relatively independent of temperament.

The methods applied in this study allow the possibility of further rotation of the value factors to achieve more satisfactory simple structure, based upon the augmented factor matrix (Table 8 and Table 4). The graphical plots of each of the five factors against each other, based upon all twenty variables, do not suggest that the structure would be improved with additional rotations. Thus, the factor solution based upon only the thirteen ways of life is not modified by the addition of factor loadings upon the seven temperament variables.

The results of this study strongly suggest substantial relationships between the domain of temperament and that of value, especially

for values represented by Factors B and C, Enjoyment and Progress in Action, and Withdrawal and Self-Sufficiency. Relations are less marked between temperament areas and value Factor D, Receptivity, and are much less marked between temperament areas and value Factors, A and E, Social Restraint, and Self-Indulgence.

It does not follow, of course, that no temperament or personality traits are operative in the favor shown by individuals to the value Factors A and E. Temperament traits have, in this study, been measured only by the *Thurstone Temperament Schedule*; it may well be that other instruments would reveal further relations to the values isolated by *Ways to Live*, or to values as isolated by other instruments. Even if such further relations were found, however, their discovery would not indicate that temperamental traits are the sole determinants of value choices. Biological differences among individuals may serve as determinants; value patterns dominant in the culture in which choices are made may influence such choices; the particular social roles of an individual may contribute their influence.[4]

The limited purpose of the present study was to explore the nature of relationships between temperamental traits and the selection of values, and to illustrate the availability of factor-analytic methods for the study of such relations between temperamental and valuational domains.

REFERENCES

1. DWYER, P. S. The determination of the factor loadings of a given test from the known factor loadings of other tests. *Psychometrika*, 1937, **2**, 173–178.
2. MORRIS, C., & JONES, L. V. Value scales and dimensions. *J. abnorm. soc. Psychol.*, 1955, **51**, 523–535.
3. MOSIER, C. I. A note on Dwyer: The determination of the factor loadings of a given test. *Psychometrika*, 1938, **3**, 297–299.
4. THURSTONE, L. L. *Multiple-factor analysis*. Chicago: Univer. of Chicago Press, 1947.
5. THURSTONE, L. L. *The Thurstone Temperament Schedule*. Chicago: Science Research Associates, 1950.
6. ZIMMERMAN, W. S. A simple graphical method for orthogonal rotation of axes. *Psychometrika*, 1946, **11**, 51–56.

Received January 12, 1956.

[4] A study of biological and social determinants of value is included in the book *Varieties of Human Value*, by Charles Morris, published in 1956 by the University of Chicago Press. Some use is also made in this book of Cattell's *16 Personality Factor Test* and Allport and Vernon's *Study of Values*.

ENHANCEMENT OF CONTRAST AS COMPOSITE HABIT[1]

DONALD T. CAMPBELL
Northwestern University

One important aspect of the general syndrome of social stereotyping is *enhancement of contrast* or the exaggeration of relative differences between social groups. Thus if on intelligence tests in New York city schools Jewish students test slightly higher than white Christian students, and Negro students slightly lower, these small differences are exaggerated in social stereotypes about the students into the judgments that Negroes are "dumb" and Jews are "smart." The perceived direction of difference is in agreement with the facts, but the proportional difference is exaggerated, the contrast enhanced, and, of course, unwarranted genetic causation imputed.

Objective studies demonstrating the effect are few in number, for such demonstration requires social judgments in situations where the appropriate facts are known, and in which a language of social judgment commensurate with the social facts can be employed. A slight degree of such bias has recently been demonstrated for white students' estimates of the examination scores of their Negro and white classmates (3). Kephart (5) has demonstrated a quite marked tendency for policemen to exaggerate the proportion of the crimes in their own districts committed by Negroes, and by inference this implies an exaggeration of the Negro-white differential in crime rates. The other available data are somewhat tangential to the usual discussions of stereotyping. In the Hartshorne and May (4) studies of character, it was possible to compare boys and girls on both objective tests of character and on reputed character as judged by classmates. They suggest "prejudice" of the form here called "enhancement of contrast" as the explanation of instances in which objective tests showed slight sex differences and reputation tests showed large ones. Thus on objective tests of "service," girls scored slightly higher than did boys (*CR* 1.9), while in reputation the difference was in the same direction but was much larger (*CR* 7.9). For "persistence" the same bias in reputational judgments was obtained (*CR*s 1.7 and 7.6). But for "inhibition" the sex difference appeared equally in objective tests and reputation (*CR*s 5.5 and 5.0). The comparison was not made for measures of "deceit."

While the empirical evidence is limited, it may be judged in sufficient accord with many less formal observations to establish enhancement of contrast as a frequent characteristic of social judgments. In specific instances this bias might be attributed to the effects of traditional attitudes transmitted from generation to generation—a learned mistaken generalization, biasing judgments of specific individuals in the concrete situation. Although this explanation remains unrefuted for the data cited, it is the effort of this paper to derive such a tendency toward bias from more general cognitive principles that would generate it even in the absence of biased traditional attitudes.

Enhancement of contrast seems to be anticipated by several approaches to cognitive functioning. In the philosophical criticism of common thought processes it appears under such labels as the tendency toward thinking in all-or-none, black-or-white, terms. Among the stubbornly recurrent errors of inference found in the teaching of logic is the error of the undistributed middle term. This error can be interpreted psychologically as a tendency to treat the pairs of terms as either totally overlapping or totally different, failing to note the possibility of partial overlap. The phraseology itself comes from the language of perception, and is related to the classification of optical illusions into errors of contrast and errors of assimilation. The view is compatible with the general spirit of Gestalt psychology, although a specific derivation might be difficult to achieve.

In calling attention to this bias, or to any other aspect of stereotyping, the emphasis of

[1] This study was initiated while the author was a visiting associate professor at Yale University, Spring, 1955. It has been supported by the Yale program of research on attitudes and communication which is financed by a grant from the Rockefeller Foundation and which is under the direction of Carl I. Hovland. Lyman W. Porter, Sarah Kendrick Martin, and Curtis E. Thomsen have contributed to the design and execution of the study.

the social psychologist is not only upon the fact of cognitive imperfection, but upon the *kind* of imperfection. The emphasis is on the constant or systematic errors. Learning theory, with its primary focus on the *degree* of perfection of the cognitive achievement rather than on the *kind* of imperfection, might seem less likely to be related to the principle. Yet there are treatments of systematic error to be found in the learning literature. A most pervasive principle predicting biased responses is that of stimulus generalization or stimulus confusion on the basis of stimulus similarity. Applied to the judgment of individuals in situations of group difference, the principle of stimulus generalization seems to predict an inevitable enhancement of contrast at stages of imperfect learning.

Consider as a paired associates task the problem of a student learning about the classroom competence of each of his Negro and white classmates, in a situation where on the average the white students do better. If skin color provides a significant degree of stimulus similarity, then stimulus generalization will lead to the confusion of one white student with other white students, one Negro student with other Negro students, prior to the completion of learning. The net result will be a tendency to confuse a given white student with the other white students, and thus to distort the judgments of individual white students in the direction of the average white student, and correspondingly for the Negro students. This distortion will be greatest for those students deviating most from their similarity group, including the least competent whites and the most competent Negroes, resulting in an underestimation of the intergroup overlap. (A more detailed derivation is provided below, in terms of the experimental analogue.)

This derivation seems to place the enhancement of contrast firmly within the range of phenomena covered by learning theory. However, it involves an extension of theory into a setting not yet studied. The interest of the social psychologist is often in the cumulative effects of a large number of "habits" or "associations" achieved contemporaneously, and not carried to perfection. This is a problem of multiple habit combination, focused not on the degree of achievement of the elementary association units, but rather on the collective resultant response tendencies. The term *composite habit* is used to refer to this problem. In our hypothetical classroom setting there is interest in the composite habit resulting from the imperfect learning of a number of paired associations of students and classroom performances. Because of the remoteness of the derivation from any actual learning experiment, it seemed desirable to check it through an appropriate laboratory task. The study reported below was conducted for this purpose.

METHOD

The apparatus consisted of a board 8½ × 26 in. with raised edges to hold in place a sliding aluminum cover. There were two apertures in this cover, in horizontal alignment with each other. The smaller one (1 × 1.5 cm.) at the left exposed a nonsense syllable. To the right of this was a long rectangular opening (1 × 12.5 cm.) through which only blank white paper could be seen. By searching with his pencil in this rectangular area, the *S* could find a vertical slot, cut in an underneath cardboard key, but not visible because covered with a record sheet, a carbon, and the top piece of blank paper. The *S*'s task was to learn the correct location of the slot for each syllable. The *S* made an initial guess (which he circled to make the record clear) and then starting from the initial guess searched the area for the slot in such a fashion as to minimize the length of the search path. When the slot was found, he moved the aluminum cover one step, exposing the next syllable and search space, and so on through the 20 items on the stencil. At this point the *E* quickly removed the stencil sheet assembly and inserted a new one, containing the same 20 items in a different order. The stencil sheet and the covering sheets were stapled together to insure easy and accurate placement. Eight random item arrangements were used, and these repeated in random orders until learning reached the criterion.

In selecting the groups of syllables there was an effort to provide two definite groups in terms of similarity, without the bases of grouping being obvious to the *S*s. One similarity group consisted of seven syllables all having a middle *E* (QEB, GEP, ZER, BEH, FEQ, KEV, PEZ); the other consisted of seven syllables all having a terminal *X* (CUX, DAX, ZOX, BIX, GUX, KAX, RIX). To further reduce the likelihood that the *S*s perceive the classification and thus treat the assignment as a task for concept formation or problem solving, there was a third group of syllables having low similarity both to one another and to either of the main groups (FUB, VAF, QIW, RUC, FOJ, LIH). As far as could be ascertained, none of the *S*s discovered the similarity groupings or the purpose of the experiment.

Centered within the 12.5 cm. search space a distance of 5.5 cm. was selected for the placement of the vertical slots. This distance was divided into 11 loci, .5 cm. apart, and numbered for designation purposes from left to right. The seven —E— syllables were associated with positions 1 through 7, in the order listed above. The seven — —X syllables were associated respectively with positions 5 through 11. The 6 buffer syllables were assigned positions 1, 3, 5, 7, 9, and 11, respectively. The

numbering and spacing of the loci were not visible to the *S*s, and the smooth surface of the aluminum cover provided no specific position cues.

The *S*s, who were paid, were recruited primarily from introductory psychology courses. The first group of eight worked in two-hour sessions, up to a total of eight hours, or until the learning criterion was reached. The next twelve worked one hour at a time for up to five hours. The criterion at which learning was stopped was an average of one spacing unit (.5 cm.) of error per item for five pages or stencils. The speed of working and rate of learning varied widely. The *S*s took from 1 to 6 hours to reach the criterion, and from 15 to 115 trials, or stencils.

THE PREDICTION

In a nonsense syllable experiment, the troublesome problem of defining stimulus similarity by stimulus characteristics can be solved in part by defining similarity in terms of identical elements. To provide a detailed illustration the character of the prediction from stimulus generalization, the following quantitative assumptions have been made: Each shared letter shall provide 1 unit of similarity; a shared letter in the same position in the syllable 2 units; a common two-letter sequence in the same position in the syllable 2 additional units. Under this scoring the "similarity" of a syllable with itself is 10 units. The similarity with another syllable from its similarity group is at least 2 units. Table 1 has been prepared on this basis, showing for each syllable the similarity units associating it with each one of the response positions, after at least one trial. Note the irregularities introduced through extraneous incidental similarities, including similarities with the buffer syllables.

This distribution of values can be taken as indicating the tendency of one syllable to be confused with the various others, or to be responded to as though it were one of the others. It can be taken as a plot of the response strengths (elicited through stimulus similarity) for each one of the response positions. The average resultant response to a given syllable is assumed to be some kind of an average of the competing response tendencies. Simple means of the distributions have been computed and these are presented graphically in Fig. 1. This averaging can be envisaged as occurring in two different ways, although the present analysis does not discriminate between them. On the one hand, each individual response might represent a spatial compromise. On the other hand, over a population of trials and persons, each position may occur in proportion to its strength, thus giving an average of responses the compromise character. The assumption of response combination or compromise is probably not useful for many learning situations, but does seem appropriate for that class of response called "judgments."

TABLE 1

PREDICTED DISTRIBUTION OF STRENGTHS OF EACH POSITIONAL RESPONSE FOR EACH SYLLABLE ON THE BASIS OF STIMULUS GENERALIZATION STATED IN TERMS OF IDENTICAL ELEMENTS

		Response Positions											
		1	2	3	4	5	6	7	8	9	10	11	Mean
True Positions	Buffer Syllables	FUB		VAF		QIW		RUC		FOJ		LIH	
	—E—'s	QEB	GEP	ZER	BEH	FEQ	KEV	PEZ					
	— —X's					CUX	DAX	ZOX	BIX	GUX	KAX	RIX	
Response Strengths for:													
QEB		12	2	2	3	5	2	2	1				3.21
GEP		2	10	2	2	2	2	3		2			3.88
ZER		2	2	10	2	2	2	5				1	4.41
BEH		4	2	2	10	2	2	2	2			2	4.57
FEQ		5	2	3	2	11	2	2		2			4.31
KEV		2	2	3	2	2	10	2			2		5.08
PEZ		2	3	3	2	2	2	11					4.96
CUX		2				10	2	5	2	6	2	2	6.81
DAX				2		2	10	2	2	2	6	2	7.36
ZOX				2		2	2	11	2	4	2	2	7.37
BIX		2			2	4	2	2	10	2	2	8	7.71
GUX		2	2			6	2	4	2	10	2	2	7.00
KAX				2		2	8	2	2	2	10	2	7.80
RIX				1		4	2	4	6	2	2	12	8.52

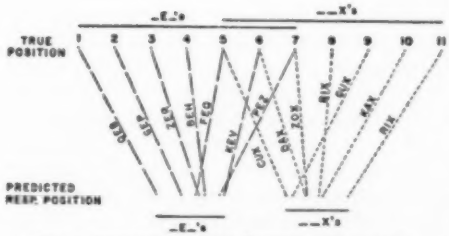

FIG. 1. PREDICTED MEAN RESPONSE POSITIONS

In Fig. 1 can be seen the predicted distortion of the extreme members of each similarity group toward the group mean, and the resulting reduction of overlap between the two groups.

RESULTS

Each trial, or sheet of 20 items, was scored in two ways, paying attention only to the initial responses to each item. For the degree of learning, the number of .5 cm. units of error for each of the 20 items was summated. For a measure of enhancement of contrast, attention was paid only to the syllables of the overlap region, positions 5, 6, 7, syllables FEQ, KEV, and PEZ; CUX, DAX, and ZOX. The sum of the initial guesses, scored in terms of the basic eleven steps appropriately extended, for the three _E_ syllables was subtracted from the sum of the initial guesses for the three __X syllables. In terms of the correct positions, the two sets of 3 syllables were equivalent. Random errors around the correct positions would result in differences with a mean of zero. The expected bias would distort _E_ syllables to the left, __X syllables to the right, resulting in a positive difference score. This bias, being based on generalization of learned associations, was expected only after some degree of learning had taken place, and of course would disappear as learning approached perfection.

The first trial error scores ranged from 74 to 177, and some *S*s showed initial increases in errors and long plateaus before any improvement. An error score of 60 was selected as a suitable point to take as evidence of some learning. All bias scores were summated from that point to the point where the error score averaged 20 or below. Such mid-learning bias scores were computed on all 20 *S*s. For 19 the sum was positive, indicating a net error in the enhancement of contrast direction. This outcome provides a one-tailed p value of $< .00002$, confirming the prediction with a high degree of significance. While the bias could not be expected to show before some learning has taken place, it is hard to designate what constitutes a "chance" score on this particular task, and hence what indicates the beginning of improvement. For this reason, the selection of the initial point of 60 error units per trial may seem somewhat arbitrary. If bias scores are summed for all trials, including the first one, up to the criterion for learning, the score is positive for the 16 out of 20 *S*s, significant at the $< .006$ level. The few additional exceptions that occur seem explainable in terms of random but large "mistaken notions" about certain syllables in the prelearning plateau.

Figure 2 shows the average trend in the bias score over the course of learning. This is a Vincent-type curve, achieved by computing for each *S* separately the average score for five equal-trial steps between his 60 and 20 error points. These five values from each *S* were then averaged. Figure 2 confirms the expectation that the bias should first rise as learning starts and then fall as learning approaches perfection.

Figure 3 shows for the second stage of learn-

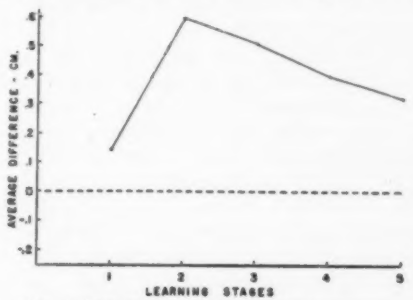

FIG. 2. DEGREE OF CONTRAST ENHANCEMENT BY STAGE OF LEARNING

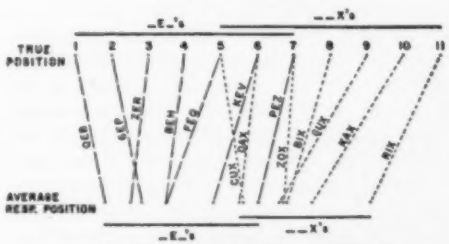

FIG. 3. AVERAGE RESPONSE POSITIONS AT LEARNING STAGE 2

ing the average response position for each of the syllables in the two similarity groups, and thus gives a graphic, over-all picture of the "stereotyping." In an experiment such as this, there is a problem of sampling syllables. The detailed interpretation of the present data is weakened by the fact that only this one set of syllables was used, and that a specifie syllable was assigned to the same position for all *S*s. Differences in the rates of learning specific to each syllable could affect the apparent distortion shown. With these qualifications in mind, Fig. 3 can be examined for general conformity to further implications of the stimulus generalization principle. Note in Fig. 1 that stimulus generalization *does not* predict that the *absolute* mean difference between the two groups as a whole will be exaggerated. Rather, members of each group are drawn toward their own group mean. And since some similarity is shared between the two groups, the group means will tend to be drawn together. These predictions seem roughly confirmed. Further, not only will the individuals in the overlap positions be distorted, but also those which deviate from their group mean at the non-overlap extremes. Because of the similarity between the two groups, these outward extremes should be even more distorted than the overlapping extremes. This prediction seems roughly confirmed, although obscured by a general leftward bias. These predictions and findings correspond in part to the social-psychological notion that stereotyping involves both reduction of overlap between groups and underestimation of variation within groups.

In this setting, reduction of overlap is not synonymous with exaggeration of the average absolute difference between groups. It is rather the exaggeration of the relative difference—relative to the over-all range of judgments made. This point might be a source of confusion in comparing the present predictions and findings with the study of children's estimates of classmates' examination grades (3). In that study, the results were reported as an exaggeration of differences between groups, without making a distinction between relative and absolute differences. Relative differences were at issue, however, as the judgments and test scores were independently converted into standard scores in terms of their own distributions. Were the distribution of average response positions in Fig. 3 to be transposed into standard scores and the mean difference between the __E__'s and the __ __X's to be computed in terms of these standard scores, a comparable exaggeration of group differences would be found. A similar treatment is provided in the Hartshorne and May data cited (4), through the use of critical ratios to measure overlap. In natural situations, it will be very difficult to use absolute comparisons with confidence, and this will limit the kind of predictions which can be tested. For example, it would be impossible with standard scores to test for the underestimation of differences within groups separately from the exaggeration of differences between groups.

DISCUSSION

The derivation of the bias of enhancement of contrast from the principle of stimulus generalization seems justified as an extension of the facts of learning and by the illustrative experiment here reported. This approach makes the bias a natural, automatic, and inevitable aspect of imperfect learning about the individual members of overlapping groups. To say this does not remove the onus of bias nor correct the injustices that derive from it.

The derivation, like others based upon cognitive processes, leads to the prediction of the enhancement of contrast on the part of all observers of the situation, including the Negro students as well as the white ones in the illustration we have employed. It must be noted that this expectation is not confirmed in the one relevant study available (3), where the Negro-white differential in exam grades was exaggerated by the white students but not by the Negro students. The total cumulative research lying behind the stimulus generalization principle is so great that one must expect the principle to be operating in this situation, although possibly in conjunction with other sources of distortion which counteract it. More extensive data are needed on this point.

At the present time in social psychology the products of experience are termed *perceptions* rather than *habits* or *response tendencies*, particularly where the products are imperfect or biased. And while introductory psychology, experimental psychology, and learning theory have become more and more dominated by the

behavioristic stimulus-response language, social and personality psychology have become more dominated by subjective, perceptual, phenomenological, or conscious-experience terminologies. Upon careful inspection it often turns out that these languages are largely interchangeable, are expressing the same kinds of environmental-behavioral regularities, are making the same kinds of predictions. But this interchangeability is so little noted that the semantic confusion remains a big problem for an integrated psychology. In the initial discussion we started out with "judgments of" persons. These are usually interpreted as "perceptions of" persons, but our main discussion has been in terms of "responses to" persons. While this led to a different choice of theoretical models, the phenomena covered have been no different. Thus although the lower line in Fig. 3 has been labeled "average response position," it could have been labeled "perceived position" with equal appropriateness. In dealing with the confusion of stimuli or distortion through stimulus generalization, the behaviorist tends to use language implying the distortion of response to an unchanging stimulus. The subjective perceptionist tends to use language implying distortion in the perceived object. While in specific instances like the present one the intersubstitutability of languages may seem obvious, the superficial dissimilarities are often so great that well-established learning phenomena may seem new when rediscovered in perceptual terms (2, pp. 167, 168). Were all perceptionists as straightforward as Asch (1, pp. 158–59) in positing an isomorphism of subjective experience and action (and even he has done so explicitly for but a limited range of conditions) the identity of the formulations would become very patent indeed. Focusing attention in learning studies on the *actual resultant responses* rather than upon the degree of perfection of responses may help make the isomorphism of perception and response more obvious.

SUMMARY

Enhancement of contrast is noted as a bias occurring in judgments made of individuals belonging to two overlapping groups. This bias is derived from the learning theory principle of stimulus generalization as a function of similarity. An experimental analogue is tested, in which two similarity groups of nonsense syllables are associated with a partially overlapping series of loci on a spatial continuum. At stages of incomplete learning the predicted enhancement of contrast or exaggeration of difference between the two similarity groups of syllables is found. The principle is discussed in terms of social stereotypes, and as an illustration of the equivalence of perceptual and behavioral languages for describing the results of experience.

REFERENCES

1. ASCH, S. E. *Social psychology.* New York: Prentice-Hall, 1952.
2. CAMPBELL, D. T. Operational delineation of "What is learned" via the transposition experiment. *Psychol. Rev.*, 1954, **61**, 167–174.
3. CLARKE, R. B., & CAMPBELL, D. T. A demonstration of bias in estimates of Negro ability. *J. abnorm. soc. Psychol.*, 1955, **51**, 585–588.
4. HARTSHORNE, H., & MAY, M. A. *Studies in the nature of character:* I. *Studies in deceit.* II. *Studies in service and self-control.* New York: Macmillan, 1928, 1929.
5. KEPHART, W. M. Negro visibility. *Amer. sociol. Rev.*, 1954, **19**, 462–467.

Received January 18, 1956.

A FACTORIAL STUDY OF DOGMATISM AND RELATED CONCEPTS

MILTON ROKEACH[1] AND BENJAMIN FRUCHTER
Michigan State University *The University of Texas*

The major purpose of this study is to determine whether the concept of dogmatism (6), and its operational measurement by means of the Dogmatism scale (8), has content, as determined from its analysis in a battery with related scales, which distinguishes it from other well-known concepts such as authoritarianism, ethnocentrism, and rigidity. Elsewhere the concept of dogmatism has been given the following threefold definition: it is "(*a*) a relatively closed cognitive organization of beliefs and disbeliefs about reality, (*b*) organized around a central set of beliefs about absolute authority which, in turn, (*c*) provides a framework for patterns of intolerance and qualified tolerance toward others" (6, p. 195). In line with this definition, the Dogmatism scale is conceived to measure both *general* authoritarianism and *general* intolerance which is not restricted to any specific ideological content. In contrast, the F scale (1) is assumed to measure primarily right authoritarianism rather than general authoritarianism and the E scale (1) to measure rightist forms of intolerance rather than general intolerance.

From dogmatism thus defined follows a distinction between dogmatism and rigidity. While dogmatism refers to total *systems* of beliefs and disbeliefs which are closed or resist change, rigidity refers only to single beliefs, sets, or expectancies which resist change. This distinction has received some experimental support in a study by Rokeach, McGovney, and Denny (7). In that study subjects were presented with a cognitive task the solution to which required the overcoming of three separate sets and then their integration. Subjects scoring high on the Gough-Sanford Rigidity Scale (4) were significantly slower than those scoring low in overcoming each of the three sets. On the other hand, subjects scoring high on the Dogmatism scale were significantly slower than those scoring low in integrating the three new sets into the problem solution after the three older sets had been overcome.

[1] This study was carried out while the senior author was a Faculty Research Fellow of the Social Science Research Council, 1951–1954.

The purpose of this study is not to demonstrate that dogmatism is a factorially pure measure independent of other measures or factors. Rather, its purpose is to determine empirically whether dogmatism as conceived and measured can be discriminated factorially, according to the considerations presented above, from the variables measured by the F scale (1), the E scale (1), and the Rigidity scale (4).

A secondary purpose of the present study has to do with the composition of the Dogmatism scale itself. The form employed in this study, as shown in Table 1, is composed of 43 items. Several of these items attempted to get at certain functional aspects of dogmatism. Thus, five items (Items 33–37) have to do with attitudes of acceptance or rejection of the self, and eight others (Items 36–43) have to do with a paranoid outlook on life. The question arises, are we justified in assuming that the self-rejection items, the paranoia items, and the other items more directly related to dogmatism belong together in the same scale? To investigate this question these three sets of items were scored separately and treated as independent variables in the analysis.

Finally, it was hypothesized that ideological dogmatism has its major motivational basis in anxiety. To the extent that this hypothesis is valid the three parts of the Dogmatism scale together with an Anxiety scale should emerge to define one of the factors in the analysis.

METHOD

A battery of ten scales was administered to 207 subjects (*Ss*) attending college in the metropolitan New York area. The scales were administered in group sessions under anonymous conditions. The *Ss* were students in elementary psychology courses. The items from the various scales were placed in random order in a single instrument and the *Ss* responded to each item on a 7-point agree-disagree scale from which the middle point had been excluded in order to avoid the piling up of responses at neutrality. For a more detailed description of the composition of the present sample and the over-all research context within which the present study is placed, see Rokeach (8).

The following scales were administered:

1. *Anxiety Scale.* Composed of 30 MMPI items found by Welch (9) to measure anxiety (e.g., "I have nightmares every few nights").

TABLE 1

THE DOGMATISM SCALE

1. A man who does not believe in some great cause has not really lived.
2. My blood boils whenever a person stubbornly refuses to admit he's wrong.
3. The worst crime a person could commit is to attack publicly the people who believe in the same thing he does.
4. It is when a person devotes himself to an ideal or cause that he becomes important.
5. Man on his own is a helpless and miserable creature.
6. In times like these, a person must be pretty selfish if he considers primarily his own personal happiness.
7. In a discussion people have a way of getting back at you by accusing you of quibbling and hair-splitting.
8. Communism and Catholicism have nothing in common.
9. There is nothing new under the sun.
10. To compromise with our political opponents is dangerous because it usually leads to the betrayal of our own side.
11. While I don't like to admit this even to myself, I sometimes have the ambition to become a great man, like Einstein, or Beethoven, or Shakespeare.
12. Unfortunately, a good many people with whom I have discussed important social and moral problems don't really understand what's going on.
13. It is sometimes necessary to resort to force to advance an ideal one strongly believes in.
14. It is better to be a dead lion than to be a live dog.
15. When it comes to differences of opinion in religion we must be careful not to compromise with those who believe differently from the way we do.
16. If a man is to accomplish his mission in life it is sometimes necessary to gamble "all or nothing at all".
17. A person who thinks primarily of his own happiness is beneath contempt.
18. It is better to be a dead hero than a live coward.
19. Nothing is impossible.
20. If I had to choose between happiness and greatness, I'd choose greatness.
21. I am not likely to speak to people until they speak to me.
22. To compromise with our political opponents is to be guilty of appeasement.
23. It's all too true that most people just won't practice what they preach.
24. Once I get wound up in a heated discussion I just can't stop.
25. The United States and Russia have just about nothing in common.
26. In a heated discussion I generally become so absorbed in what I am going to say that I forget to listen to what the others are saying.
27. To one who really takes the trouble to understand the world he lives in, it's a relatively easy matter to predict future events.
28. To achieve the happiness of mankind in the future it is sometimes necessary to put up with injustices in the present.
29. The present is all too often full of unhappiness. It is the future that counts.
30. It is by returning to our forgotten and glorious past that real social progress can be achieved.
31. I am afraid of people who want to find out what I'm really like, for fear they'd be disappointed in me.
32. I feel self-conscious when I'm with people who have a superior position to mine in school or business.
33. I'd like it if I could find someone who would tell me how to solve my personal problems.
34. My hardest battles are with myself.
35. At times I think I am no good at all.
36. All too many people are failures and it is the system which is responsible for this.
37. It is only natural for a person to have a guilty conscience.
38. It is generally safer to trust nobody.
39. Much of the time I feel as if I have done something wrong or evil.
40. I have often felt that strangers were looking at me critically.
41. If given the chance I would do some things that would be of great benefit to the world.
42. I have often felt that people say insulting and vulgar things about me.
43. I am sure I am being talked about.

2. *Paranoia Subscale.* Composed of Items 36–43 in Table 1 (e.g., "It is safer to trust nobody".[2]

3. *Self-Rejection Subscale.* Composed of Items 31–35 in Table 1 ("I am afraid of people who want to find out what I'm really like, for fear they'd be disappointed in me").[3]

4. *Dogmatism Subscale.* Composed of the first 30 items in Table 1 (e.g., "A man who does not believe in some great cause has not really lived").[4]

5. *Authoritarianism (F Scale).* Composed of 29 items (e.g., "Obedience and respect for authority are the most important virtues children should learn") (1, Form 45-40, pp. 255–257).

6. *Rigidity Scale.* Composed of 22 items (e.g., "I wish people would be more definite about things") (4).[5]

[2] Items 38–42 are taken from the MMPI (4).

[3] Items 33, 34, 35 are taken from Berger (2), Items 36 and 37 from MMPI (4).

[4] The corrected reliability of this scale is .75. Since this study was completed the Dogmatism scale has undergone further revision, primarily for the purpose of increasing its reliability. A slightly modified version of this scale was employed in a study by Rokeach, McGovney, and Denny (7). The final form of the Dogmatism scale is composed of 40 items and is reported fully in a separate publication (8).

[5] The writers are grateful to Drs. Gough and Sanford for their courtesy in permitting use of this scale. This scale is also included in Gough's California Psychological Inventory (copyright 1951). Some sample

TABLE 2
CORRELATION MATRIX
($N = 207$)

	Variable										
	1	2	3	4	5	6	7	8	9	10	r^*_{tt}
1. Anxiety		.65	.71	.33	.17	.14	.13	−.13	.22	−.05	.90
2. Paranoia			.52	.41	.30	.29	.20	−.08	.31	−.03	.60
3. Self-rejection				.30	.14	.07	.10	−.14	.11	−.02	.64
4. Dogmatism					.64	.62	.52	.23	.12	.33	.75
5. Authoritarianism (F Scale)						.69	.63	.43	−.08	.56	.84
6. Rigidity							.54	.35	.02	.40	.83
7. Ethnocentrism (E Scale)								.44	−.22	.62	.72
8. Conservatism (P.E.C.)									−.39	.60	.62
9. Left Opinionation										−.51	.74
10. Right Opinionation											.77

* Odd-even reliability estimates of the scales corrected for length.

TABLE 3
FACTOR LOADINGS

Variable	Rotated Loadings			Centroid Loadings			
	I	II	III	I	II	III	h^2
1. Anxiety	.77	−.25	.27	.442	−.708	.173	.727
2. Paranoia	.72	−.14	.26	.468	−.600	.134	.597
3. Self-rejection	.69	−.29	.37	.412	−.656	.313	.698
4. Dogmatism	.46	.21	.62	.708	−.139	.341	.637
5. Authoritarianism (F scale)	.27	.48	.66	.784	.185	.297	.737
6. Rigidity	.23	.32	.71	.680	.123	.421	.652
7. Ethnocentrism (E scale)	.21	.59	.47	.728	.273	.099	.614
8. Conservatism (P.E.C.)	−.07	.69	.19	.487	.519	−.129	.523
9. Left opinionation	.17	−.63	.26	−.169	−.500	.468	.498
10. Right opinionation	.03	.85	.19	.642	.549	−.227	.765

7. *Ethnocentrism (E Scale).* Composed of 10 items about Negroes and other minority groups, but excluding items about Jews. (e.g., "Manual labor and unskilled jobs seem to fit the negro mentality and ability better than more skilled or responsible work") (1, Item 32, p. 128; Items B1–B4, and C1–C5, p. 142).

8. *Political-Economic Conservatism (PEC Scale).* Composed of five items (e.g., "In general, full economic security is bad; most men wouldn't work if they didn't need the money for eating and living.") (1, Form 45-40, p. 169).[6]

9. *Left Opinionation Scale.* Composed of 20 items designed to measure degree of intolerance in a left direction (e.g., "It's just plain stupid to keep Red China out of the United Nations").[7]

10. *Right Opinionation Scale.* Composed of 20 items designed to measure degree of intolerance in a right direction (e.g., "It's the agitators and left wingers who are trying to get Red China into the United Nations").

RESULTS AND DISCUSSION

The scores derived from the several scales were intercorrelated by the Pearson product-moment method and the resulting coefficients as well as the corrected odd-even reliability estimates are presented in Table 2. The correlation matrix was factored by the centroid

items are: "I wish people would be more definite about things;" "Once I have made up my mind I seldom change it;" "I always see to it that my work is carefully planned and organized;" "I like to have a place for everything and everything in its place."

[6] However, the P.E.C. item "Most government controls over business should be continued even though the war is over" was revised to read: "Most government controls over business should be continued and even strengthened."

[7] A fuller discussion of the total Opinionation scale and the several subscales thereof is presented elsewhere (8). Our purpose for including the Left and Right Opinionation subscales in this study was to find out whether these measures, along with the P.E.C. scale, would emerge together on a liberalism-conservatism factor.

method and the three factors extracted were rotated to approximate orthogonal simple structure (3). The rotated loadings, centroid loadings, and communalities for the ten variables are shown in Table 3.

The variables with the highest loadings on the first rotated factor are the Anxiety, Paranoia, and Self-Rejection. Dogmatism also has a substantial loading on this factor. Left and Right Opinionation and Political-Economic Conservatism have the highest loadings on the second rotated factor. The E scale and the F scale also have substantial loadings toward the "right" end of the bipolar factor. With respect to the third rotated factor it is seen that Rigidity, the F scale, and Dogmatism have the highest loadings and the E scale has a moderate loading on it. While it is difficult to attach definite labels to these three factors, Factor I may be tentatively identified as representing anxiety, Factor II liberalism-conservatism, and Factor III rigidity-authoritarianism.

More important for the purposes of this study than the clear-cut identification of these factors is any assistance they may be in clarifying the issues stated at the beginning of the paper.

1. *Is the Dogmatism subscale discriminable from the F scale, the Rigidity scale and the E scale?* It is seen that while both dogmatism and authoritarianism have their major loadings on Factor III, the pattern of loadings for these two variables differs on the first two factors. The F scale has a substantial loading toward the right end of the liberalism-conservatism factor (Factor II) and a negligible loading on the anxiety factor (Factor I) in contrast to the Dogmatism subscale which has the opposite profile on these two factors. These findings suggest that even though Dogmatism and F are correlated with each other ($r = .64$) Dogmatism is relatively independent of the left-right continuum while the F scale is not. The situation for the two scales is the reverse with respect to the anxiety factor, with dogmatism showing a substantially higher loading than authoritarianism on that factor.

The results bearing on the Rigidity scale indicate that it has its major loading on Factor III. It has minor loadings on the first two factors and comes closest to being a pure representative of the third factor. As for the E scale, it has its largest loading on Factor II (liberalism-conservatism) and also a moderately high loading of Factor III (rigidity-authoritarianism).

It is thus seen that Dogmatism differs from the E and F scales in that it is independent of the left-right continuum and has a strong component on the anxiety factor, whereas the other two scales are oriented toward the right end of the liberalism-conservatism factor and have only negligible loadings on the anxiety factor. Dogmatism can be distinguished factorially from rigidity in that the latter has its only substantial loading on the rigidity-authoritarianism factor.

2. *Are dogmatism, self-rejection, and paranoia factorially similar?* It is seen in Table 3 that these three variables do indeed have substantial loadings, along with anxiety, on Factor I. Thus, there is good supporting evidence that the eight items having to do with paranoic and the five items having to do with self-rejecting tendencies do contribute to the anxiety aspect of the total Dogmatism scale. At the same time it is noted that there is a difference between the Dogmatism subscale (30 items) on the one hand and the paranoic and self-rejection subscales on the other. The Dogmatism subscale has a relatively high loading on Factor III (rigidity-authoritarianism), while the paranoia and self-rejection variables do not.

3. *What is the relation between dogmatism and anxiety?* It is seen further that anxiety has a very heavy loading on Factor I and relatively minor loadings on Factors II and III. In accord with our hypothesis anxiety is seen to emerge on a single factor along with the three parts of the total Dogmatism scale already referred to.

SUMMARY AND CONCLUSIONS

A factor analysis was conducted using a sample of 207 New York college students, of scales representing the following ten variables: anxiety, paranoia, self-rejection, dogmatism, authoritarianism, rigidity, ethnocentrism, liberalism-conservatism, left opinionation, and right opinionation. The following three major hypotheses were investigated: (*a*) that dogmatism as conceived and measured is factorially discriminable from authoritarianism, ethnocentrism, and rigidity; (*b*) that dogmatism,

paranoia and self-rejection are factorially similar; (*c*) that dogmatism, paranoia, and self-rejection would emerge, together with anxiety, on a single factor.

On the whole it can be concluded from the data that the three hypotheses are confirmed. Factor I is composed primarily of anxiety, paranoia, and self-rejection, with dogmatism also having a substantial loading on this factor. Factor II is primarily a liberalism-conservatism factor being composed primarily of left and right opinionation, the Political-Economic Conservatism Scale, and to a lesser extent, ethnocentrism and the F scale. Factor III includes primarily rigidity, the F scale, and dogmatism.

From all the preceding we conclude that dogmatism, as conceived and represented, measures something similar to authoritarianism but independently of the left-right dimension (Factor II). It is also discriminable from rigidity, and ethnocentrism. Finally, it seems to be related to anxiety.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. BERGER, E. M. The relation between expressed acceptance of self and expressed acceptance of others. *J. abnorm. soc. Psychol.*, 1952, **47,** 778–782.
3. FRUCHTER, B. *Introduction to factor analysis*. New York: Van Nostrand, 1954.
4. GOUGH, H. G., & SANFORD, R. N. Rigidity as a psychological variable. Berkeley: Univer. of Calif. Inst. of Pers. Assessment and Res., 1952 (mimeographed).
5. HATHAWAY, S. R., & MCKINLEY, J. C. *The Minnesota Multiphasic Personality Inventory*. New York: Psychological Corp., 1943.
6. ROKEACH, M. The nature and meaning of dogmatism. *Psychol. Rev.*, 1954, **61,** 194–204.
7. ROKEACH, M., MCGOVNEY, W. C., & DENNY, M. R. A distinction between dogmatic and rigid thinking. *J. abnorm. soc. Psychol.*, 1955, **51,** 87–93.
8. ROKEACH, M. Political and religious dogmatism: An alternative to the authoritarian personality. *Psychol. Monogr.*, **70,** No. 18 (Whole No. 425).
9. WELCH, G. S. An anxiety index and an internalization ratio for the MMPI. *J. consult. Psychol.*, 1952, **16,** 65–72.

Received January 18, 1956.

PERFORMANCE AS A FUNCTION OF MOTIVE STRENGTH AND EXPECTANCY OF GOAL-ATTAINMENT

JOHN W. ATKINSON AND WALTER R. REITMAN[1]

University of Michigan

THE results of recent studies that have related thematic apperceptive measures of achievement (7) and affiliation motives (3) to performance have been consistent with a principle of performance adapted from Tolman's (11) expectancy theory: the goal-directed action tendency is a joint function of the strength of the motive and of the expectancy aroused by situation cues that performance is instrumental to attainment of the goal of the motive (2). Our purpose is to elaborate this idea as it applies to the relation of performance to measures of individual differences in human motives and to present an experimental test of a prediction when more than one motive and one expectancy are involved.

Earlier studies (1, 4, 5, 6, 12) have shown a positive relationship between performance and n-Achievement scores obtained from thematic apperceptive stories when the cues of the performance situation arouse the expectancy that a feeling of personal accomplishment will accompany a good performance. On the other hand, when the cues for performance are deliberately manipulated so that *S*s are given no reason to expect they will experience pride in accomplishment, there is no relationship between performance and n-Achievement scores (1, 4, 5). On two occasions (4, 5), the *E*'s explicit appeal for cooperation in situations not related to achievement has apparently served to engage another motive. Performance level, while unrelated to strength of n Achievement, has been found positively related to n Affiliation. French (5)[2] has also shown that when an incentive unrelated to either n Achievement or n Affiliation is offered, no systematic relationship is found between performance and either of these motives.

We conceive a *motive* as a latent disposition to strive for a particular goal-state or aim, e.g., achievement, affiliation, power. The strength of a particular motive is assessed by thematic apperception under *neutral*[3] conditions. The term *motivation* can then be used to designate the aroused state of the person that exists when a motive has been engaged by the appropriate expectancy, i.e., an expectancy that performance of some act is instrumental to attainment of the goal of that motive. If more than one of an individual's motives are engaged by expectancies that the same act will lead to several different goals, the *total motivation* for performance of that act will be the sum of the contributions made by the particular motives which have been engaged. An overdetermined action tendency, i.e., one which will serve to satisfy more than one motive, is likely to be relatively strong.

This conception implies that a positive relationship between a particular motive (e.g., n Achievement) and performance can be maximized by engaging that motive and no others in performance and can be minimized either (*a*) by failing to engage that motive at all as in (1, 4, 5) or (*b*) by systematically engaging other motives in performance *as well*, as suggested in (5, 12). In the latter case, the person who is weak in the motive the *E* may have measured may be strong in other unmeasured (ulterior) motives that are engaged in performance of the same act.

The missing link among the studies already accomplished is a demonstration that the relationship between n Achievement and performance can be substantially reduced even *when the achievement expectancy is explicitly aroused* by systematic arousal of other motives to perform the same act. This should occur, assuming an asymptote for performance, when

[1] This study was accomplished during the term of a Social Science Research Council fellowship to JWA and a National Science Foundation fellowship to WRR. We should like to acknowledge, also, assistance provided by the Center for Advanced Study in The Behavioral Sciences.

[2] French's motive scores were obtained through content analysis of responses made to a series of single-sentence descriptions of behavior which *S*s were asked to explain as a "Test of Insight." The categories of response are directly comparable to those employed in the analysis of TAT stories.

[3] A *neutral* condition is one in which no experimental attempt is made either to arouse the motive or to create an especially relaxed state prior to the writing of stories (see 7).

the total motivation of a low n-Achievement group approximates that of a high n-Achievement group. Our experimental task, then, is to engage the achievement motive *in two different situations* by instructions that are known to arouse it. But then we must manipulate other features of the two situations so that expectancies of attaining other goals are relatively weak or absent in one situation and strong in the other.

Taking a lead from Wendt (12), we designed a group test situation in which two examiners would stalk about the room with watches in hand, pacing performance with a signal every minute and, to the best of their ability, casting disapproving glances at anyone who seemed to be gazing around or not taking the task seriously. These innovations, in addition to the usual type of achievement-orienting instructions, were introduced to engage the affiliation motive, the desire to please the experimenter. Finally, in an appeal to the motivational structure of persons who might be low in both n Achievement and n Affiliation, a substantial monetary reward ($5.00) was offered to the person getting the best score on the test.

In another condition, the same achievement-orienting instruction was used to engage the achievement motive, but to minimize the arousal of other goal-expectancies, the *S*s were assigned individually to small test rooms and left to work alone without even the competitive cues provided by the sight and sound of other persons working. In this condition, the *E* was not present during performance, and there was no monetary incentive.

The performance level of persons who are high in n Achievement should be substantially higher than that of persons who are low in n Achievement in the group given achievement-orienting instructions and then left on their own to work without expectancies of satisfying other motives. The difference should be substantially reduced (or removed completely) in the other group given achievement-orienting instructions but in whom expectancies of attaining other goals (affiliation and money) were also aroused. No relationship is expected between n Affiliation and performance in either condition. In one group, only the expectancy of achievement was aroused, while in the other, expectancies of achievement and a monetary reward were also present to confound the simple relationship to n Affiliation that might otherwise have resulted from the introduction of affiliation-related cues.

METHOD

Subjects were 93 male college students selected from a group of several hundred introductory psychology students for whom measures of both achievement and affiliation motives were available.

Measurement of achievement and affiliation motives. Eight four-minute imaginative (TAT-type) stories were written under neutral (7) classroom conditions one to three months before the experimental sessions. The results of a methodological study run concurrently indicate that the first four stories in a series of eight, irrespective of pictures used, provide the greatest predicative validity for n Achievement (9). The motive scores of the present study were therefore based only on the first four stories. The rank-order score-rescore reliability was .88 for n Achievement (7) and .89 for n Affiliation (3). The distributions of scores of the larger population from which the sample was drawn were divided at the median to yield high and low groups with respect to n Achievement and n Affiliation.

Experimental conditions. Forty-nine *S*s in groups of six to ten were assigned to the *Achievement-Orientation* condition. The *E* emphasized the importance of doing well on the tasks by alluding to them as similar to tests of important abilities (executive capacity, leadership, intellectual alertness) developed by psychologists in the past. The *S*s were urged to do their best and were told, "your work will be taken as the full measure of your ability" (see 1). They were also told that they would be assigned to individual test rooms to prevent distractions.

Following this general achievement-orienting instruction, and before being assigned to individual test rooms, *S*s were given a short story to read and told to memorize its content for later recall. This task, part of another study not reported here,[4] took three minutes. It was introduced to take advantage of the aroused motivational state of the *S*s on the assumption that it would in no way affect the performance of *S*s on the subsequent tasks of the present experiment. As it turned out, this assumption was probably incorrect.

The *S*s were then assigned to individual test rooms. The first task, which consisted of drawing *X*s inside of circles, was explained to them. When the test was delivered to each *S*, he was told to begin and urged, "do your best," or, "see how good you are at this one." The *E* then noted on a stop watch the exact time the *S* started working and recorded this time in seconds immediately after leaving the room. He then allowed 10–15 seconds to elapse before delivering a test to the next *S* in an adjacent room to make it possible for him to stop each *S* exactly 14 minutes after his own starting time. The *E* entered each room and stopped each *S* within five seconds of 14 minutes after he had started.

Instructions for the second task, a set of simple

[4] Reitman, W. R., Motivation and recall of meaningful material under achievement-orientation. (Unpublished paper) University of Michigan, 1955.

three-step arithmetic problems, were then distributed and explained. The delivery of the second task followed the same procedure. Each *S* was told, "this one will demand a little more brain power," or, "this one will be a lot harder," when the test itself was delivered. The same procedure of timing *S*s was employed.

Forty-four *S*s were assigned to the *Multi-Incentive* condition. The *S*s in this condition met in groups of 10–20 and stayed together in a large classroom for performance of both tasks. The same general achievement-orienting instruction was given followed by the three-minute memorization task. The *E* then said, "Since we are interested in seeing your very best performance when you are actually putting out, we are going to award a prize of $5.00 to the person having the highest score on each of the tests. So you could stand to walk off with $10.00 for your efforts today." Finally, *S*s were told they would be asked to skip an item from time to time throughout the test period.

Throughout the performance of each test, the *E*s walked about the room, stop watches in hand, as if evaluating the seriousness of each *S*'s effort. At the end of each minute, one of the *E*s called out, "skip," at which point the *S*s stopped, skipped an item, and moved on to the next one.

Fourteen minutes were given for each task, and the monetary incentive with a reminder, "do your best," was repeated before the second task which was also paced.

The tasks. The first task consisted of a 15-page booklet containing row after row of ½ inch circles in which the *S* was instructed to draw *X*s (4). The second task, arithmetic problems, required solution of a series of three-step problems. There were 14 pages, each containing 24 problems. The operations consisted of adding or subtracting three digits, remembering the solution while adding or substracting another set of three digits, and then finally adding or substracting the two solutions (12).

RESULTS

Drawing Xs in circles. The mean number of *X*s drawn in circles was 1255.5 for the Multi-Incentive condition ($N = 44$) and 1149.1 for the Achievement-Orientation condition ($N = 51$). The difference of 106.4 between the two conditions is significant ($t = 3.62$, $p < .01$), but there was no difference between high and low n-Achievement groups within either condition.

We suspect that performance of the first task was confounded by continued attempts to practice the memory material that had been introduced at the beginning of the test period. When queried about possible effects of the memory task after the experimental session, a number of *S*s indicated that they had continued to practice recall during the first task. In light of the results of one earlier study (4) and another in preparation that show the *X*s-in-circles task to be a sensitive measure of effort and significantly related to differences in n Achievement under appropriate conditions, we assume that many of the *S*s felt it more important to concentrate on the memory material than to do their best at the relatively simple motor task that followed. Continued rehearsal of the memory material would be sufficiently distracting to influence the level of performance.

The achievement motive and arithmetic performance. The results for the second task are presented in Table 1. The arithmetic problems, which demanded continued attention and concentration, would not encourage continued attempts at memorization. We therefore view the results presented in Table 1 as a less confounded test of the initial hypothesis.

In the Achievement-Orientation condition, the performance level of the high n-Achievement group was significantly higher ($p < .01$) than that of the low n-Achievement group on both number of solutions attempted and number correct. In the Multi-Incentive condition, there was no difference between the performance levels of the two motivation groups. The difference between the differences is 19.8 for attempted solutions and 15.9 for correct solutions. Both are significant in the direction predicted (ts = 2.30 and 1.84 respectively, $p < .05$). In other words, the relationship between achievement motive and arithmetic performance was eliminated by systematically

TABLE 1

ARITHMETIC PERFORMANCE AS A FUNCTION OF ACHIEVEMENT MOTIVE AND EXPERIMENTAL CONDITION

Achievement Motive	Condition					
	Achievement Orientation			Multi-Incentive		
	N	Attempted Solutions	Correct Solutions	*N*	Attempted Solutions	Correct Solutions
High	21	M 78.1	71.6		M 67.1	60.3
		σ 24.8	24.1	24	σ 19.6	19.3
Low	30	M 60.3	55.5		M 69.1	60.1
		σ 15.7	16.4	21	σ 22.3	23.0
Diff (H-L)		17.8	16.1		−2.0	.2
σ diff.		5.93*	5.96		6.23	6.26
t		3.00	2.70		n.s.	n.s.
p		.01	.01			

* σ diffs. derived from estimate of within group variance, $df = 92$.

engaging other motives in the same performance.

An unanticipated fault in our method of pacing is chiefly responsible for the failure of over-all performance level to increase in the Multi-Incentive condition. When *E* called out, "skip," at the end of each minute, *S*s had to stop working on partially solved problems and move immediately to the next one. On the earlier task, they had merely to skip one circle. Since there were 13 such interruptions, the performance of the high n-Achievement group would suffer a decrement if it were already nearly asymptotic in the Achievement-Orientation condition. And the net gain of the low n-Achievement group, which should profit most by the arousal of other motives, would also be depressed. The high n-Achievement group actually dropped 11.0 in number of attempts and 11.3 in number correct. The low n-Achievement groups showed net gains of 8.8 in number of attempts and 4.6 in number correct. The extent to which the decrease of the high n-Achievement group may be a real decrement accompanying very intense motivation is, unfortunately, indeterminate.

Joint effect of motivation and aptitude on arithmetic performance. Lowell (6, 7) found no relationship between n Achievement and quantitative aptitude among college students. His results can be checked in the present study.

Quantitative Aptitude Scores (Q scores) from the American Council of Education Test were available for 38 *S*s in the present Achievement-Orientation condition. The data for 20 additional *S*s for whom Q scores could be obtained were added. These 20 *S*s had been run under the same Achievement-Oriented condition but as part of a methodological study (9). They were divided into high and low n-Achievement groups according to the median score of a much larger group who had taken a different and somewhat less sensitive form of the thematic apperceptive test of motivation. The obtained distribution of 58 Q scores (percentile rank using local norms) was divided into thirds to provide High, Middle, and Low aptitude groups.

The mean Q score was 64.6 for this high n-Achievement group ($N = 29$) and 57.2 for the low n-Achievement group ($N = 29$). The difference, 7.4, is insignificant ($t = 1.05$), confirming Lowell's finding.

The mean number of arithmetic problems attempted by this high n-Achievement group ($N = 29$), irrespective of aptitude, was 71.2; the mean for the low n-Achievement group ($N = 29$) was 61.7. The difference, 9.5, while less than the comparable difference in the experiment proper (Table 1) as a consequence of the less sensitive measure of achievement motive for the 20 additional *S*s, is nevertheless significant ($t = 2.06$, $p < .05$). Similarly, the performance levels of High ($N = 18$), Middle ($N = 22$), and Low ($N = 18$) aptitude groups, independent of motive strength, are significantly different as normally expected ($F = 3.23$, $df = 2$ and 55, $p < .05$).

Table 2 shows that the highly motivated groups were uniformly higher in performance

TABLE 2

ARITHMETIC PERFORMANCE AS A JOINT FUNCTION OF ACHIEVEMENT MOTIVE AND QUANTITATIVE APTITUDE (Achievement-Orientation Condition)

Quantitative Aptitude (Percentile Rank)	Achievement Motive: High			Achievement Motive: Low			Diff.	σ diff.	*t*	*p*
	N	Attempted Solutions	Correct Solutions	*N*	Attempted Solutions	Correct Solutions				
High 80–99th	11	76.9	70.5	7	72.4	68.5	4.5 2.0	8.4 8.5	n.s. n.s.	
Middle 53–75th	10	65.7	61.2	12	59.8	54.7	5.9 6.5	7.4 7.5	n.s. n.s.	
Low 3–48th	8	70.4	62.8	10	56.4	49.0	14.0 13.8	8.2 8.3	1.71 1.66	.05** .05**
Diff. (H-L)		6.5	7.7		16.0	19.6				
σ diff.		8.1*	8.2		8.6	8.7				
t		—	—		1.86	2.25				
p		—	—		.05**	.05**				

* σ diffs. derived from estimates of within group error, $df = 52$.

** One-tailed tests.

than the less motivated groups at the three levels of aptitude. But the effect of motivation was most pronounced in the group that was lowest in aptitude. Here the difference between high and low n-Achievement groups in number of attempted solutions was 14.0 ($t = 1.71$, $p < .05$ in predicted direction). It is less than half as large at the middle level of aptitude (diff. = 5.9, σ diff. = 7.4) and at the high level of aptitude (diff. = 4.5, σ diff. = 8.4). The difference between the differences of High and Low aptitude groups, 9.5, is not significant (σ diff. = 11.8). The result nevertheless suggests that persons who are low in aptitude profit most by strong motivation. The trends for number of correct solutions are very similar, as Table 2 shows, but the differences are somewhat smaller.

Table 2 shows a comparable result regarding the effect on performance of differences in aptitude, viz., the effect is greatest when motivation is low. The difference between High and Low aptitude groups also low in n Achievement was 16.0 for attempted solutions (σ diff. = 8.6, $t = 1.86$, $p < .05$ in expected direction). The comparable difference in the group who were highly motivated to achieve is only 6.5 (σ diff. = 8.1). The result is essentially the same, though in this case slightly more pronounced, for number of correct solutions. In neither case is the difference between the differences significant.

The affiliation motive and arithmetic performance. No systematic relationship between n Affiliation and performance is expected in either of the experimental conditions. Table 3 shows that none of the differences is significant when a two-tailed *t*-test is made. The tendency for the low n-Affiliation group to perform at a somewhat higher level under Achievement-Orientation can probably be attributed to the low negative correlation (−.20 to −.25) between n Achievement and n Affiliation when scores are taken from the same four stories (9). In light of this, the *positive* trend between n Affiliation and performance in the Multi-Incentive condition, particularly for correct solutions ($p = .10$), may mean that we were not completely successful in our attempt to eliminate the relationship between performance and a particular motive by systematically arousing other motives. The positive trend, while statistically very tenuous, suggests that the expectancy of pleasing the *E*s by working hard was most salient in the Multi-Incentive condition. Comparison of changes in performance between the two conditions supports this view, as *S*s in the Multi-Incentive condition high in n Affiliation and low in n Achievement ($N = 8$) averaged 16.7 more attempted solutions than *S*s with comparable motive scores in the Achievement-Orientation condition ($N = 20$); all other *S*s in the Multi-Incentive condition ($N = 37$) averaged 5.8 *fewer* attempted solutions than the comparable Achievement-Oriented *S*s ($N = 31$). The difference, 22.5, is significant (σ diff. = 10.0, $t = 2.25$ $p < .05$). The correct solutions result is very similar. The greatest increase in performance due to the presence of the *E*s and the monetary reward was among *S*s highly motivated to affiliate but weak in motivation to achieve.

TABLE 3
ARITHMETIC PERFORMANCE AS A FUNCTION OF AFFILIATION MOTIVE AND EXPERIMENTAL CONDITION

Affiliation Motive	Condition					
	Achievement Orientation			Multi-Incentive		
	N	Attempted Solutions	Correct Solutions	*N*	Attempted Solutions	Correct Solutions
High	28	M 63.3	57.7	17	M 75.1	68.3
		σ 15.8	15.9		σ 22.5	22.9
Low	23	M 73.0	67.5	28	M 66.1	57.4
		σ 26.4	25.8		σ 18.6	18.3
Diff (H-L)		−9.7	−9.8		9.0	10.9
σ diff.		6.0*	5.9		6.5	6.3
t		1.62	1.66		1.38	1.72
p		n.s.	n.s.		n.s.	.10

* σ diffs. derived from estimate of within group error, $df = 92$.

DISCUSSION

Genotype and phenotype. We interpret the results as consistent with the conception of motivation and behavior elaborated in the introductory section. The distinction between strength of particular motives (latent dispositions) and strength of motivation to perform a particular act (the sum of all aroused motives) highlights the genotype-phenotype distinction that is often made in treatments of motivation and behavior. An obvious implication of the results of this and earlier, related studies (1, 4, 5, 12) is that psychologists should stop expecting to find measures of *particular* motives and performance criteria like academic

performance, leadership, etc., to be more highly related than they have been found in the past (for example 8). From the present viewpoint, most performance criteria in real-life situations can be considered overdetermined, i.e., acts or accomplishments for which several of the individual's motives have been engaged.

The role of the situation. The present formulation calls for a good deal more thought than is commonly given the question of conceptualizing and ultimately measuring the properties of the situation in which behavior occurs. Our difficulty in attempting to create an experimental situation that would make the expectancies of achievement, affiliation, and monetary prize nearly equal in salience, as evidenced by the positive trend between n Affiliation and performance in the Multi-Incentive condition, illustrates the need for adequate methods of assessing situations in terms of the expectancies of goal-attainment they provoke. Rotter (10) and his colleagues have begun to make substantial strides in this direction.

SUMMARY

Particular motives, (Achievement, Affiliation, Power, etc.) are conceived as latent dispositions to strive for certain goal states (aims) that are engaged in performance of an act when the cues of the situation arouse the expectancy that performance of an act is instrumental to attainment of the goal of the motive. The total *motivation* to perform the act is conceived as a summation of strengths of all the *motives* that have been aroused by appropriate expectancies of goal-attainment cued-off by the situation. The relationship of achievement motive (as measured by imaginative TAT stories) to performance is shown to be significantly positive when the expectancy that performance is instrumental to producing a feeling of pride in accomplishment is aroused and few if any other expectancies of goal-attainment are aroused. When motives for other goals (affiliation and money) are *also* aroused by deliberate manipulation of situational cues that activate the *S*s' expectancies, there is no relationship between achievement motive and performance. The joint effect of quantitative aptitude and achievement motive on arithmetic performance is illustrated in a subsidiary analysis of the data.

REFERENCES

1. ATKINSON, J. W. The achievement motive and recall of interrupted and completed tasks. *J. exp. Psychol.*, 1953, **46**, 381–390.
2. ATKINSON, J. W. Explorations using imaginative thought to assess the strength of human motives. In *Nebraska symposium on motivation, 1954.* Lincoln: Univer. of Nebraska Press, 1954.
3. ATKINSON, J. W., HEYNS, R. W., & VEROFF, J. The effect of experimental arousal of the affiliation motive on thematic apperception. *J. abnorm. soc. Psychol.*, 1954, **49**, 405–410.
4. ATKINSON, J. W., & RAPHELSON, A. C. Individual differences in motivation and behavior in particular situations. *J. Pers.*, 1956, **24**, 349–363.
5. FRENCH, E. G. Some characteristics of achievement motivation. *J. exp. Psychol.*, 1955, **50**, 232–236.
6. LOWELL, E. L. The effect of need for achievement on learning and performance. *J. Psychol.*, 1952, **33**, 31–40.
7. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.
8. MORGAN, H. A. Measuring achievement motivation with "Picture Interpretations." *J. consult. Psychol.*, 1953, **17**, 289–292.
9. REITMAN, W. R., & ATKINSON, J. W. Methodological problems in the use of thematic apperceptive measures of human motivation. In J. W. Atkinson (Ed.), *The assessment of human motives.* Princeton: Van Nostrand, in press.
10. ROTTER, J. B. The role of the psychological situation in determining the direction of human behavior. In M. R. Jones (Ed.), *Nebraska symposium on motivation, 1955.* Lincoln: Univer. of Nebraska Press, 1955.
11. TOLMAN, E. C. Principles of performance. *Psychol. Rev.*, 1955, **62**, 315–326.
12. WENDT, H. W. Motivation, effort, and performance. In D. C. McClelland (Ed.), *Studies in motivation.* New York: Appleton-Century-Crofts, 1955.

Received January 20, 1956.

COGNITIVE STRUCTURE AND ATTITUDINAL AFFECT[1]

MILTON J. ROSENBERG

Yale University

UNDERSTANDING of the related processes of attitude learning and attitude change will probably be advanced by the investigation of structural relationships between attitudes and beliefs about the objects of attitudes. The present research is an attempt to verify a set of hypotheses about such relationships. These hypotheses have much in common with some that have already been presented by Cartwright (2), Hilliard (3), Smith (8), Tolman (9), and Woodruff (11, 12). However, they differ from these earlier formulations in attempting to delineate more explicitly certain variables that are assumed to covary with attitude (here defined as *relatively stable affective response to an object*). One of these variables is the intensity of a person's values. A second is the perceived importance of the attitude object in leading to or blocking the attainment of his values.

The general theoretical view underlying the present study includes the following points: (*a*) When a person has a relatively stable tendency to respond to a given object with either positive or negative affect, such a tendency is accompanied by a *cognitive structure* made up of beliefs about the potentialities of that object for attaining or blocking the realization of valued states; (*b*) the sign (positive or negative) and extremity of the affect felt toward the object are correlated with the content of its associated cognitive structure. Thus strong and stable positive affect toward a given object should be associated with beliefs to the effect that the attitude object tends to facilitate the attainment of a number of important values, while strong negative affect should be associated with beliefs to the effect that the attitude object tends to block the attainment of important values. Similarly, moderate positive or negative affects should be associated with beliefs that relate the attitude object to less important values or, if to important values, then with less confidence as to the existence of clear-cut instrumental relationships between the attitude object and the values in question.

From this view, three specific hypotheses were formulated for experimental test:

1. The degree and sign of affect aroused in an individual by an object (as reflected by the position he chooses on an attitude scale) vary as a function of the algebraic sum of the products obtained by multiplying the rated importance of each value associated with that object by the rated potency of the object for achieving or blocking the realization of that value.

2. The degree and sign of affect aroused in an individual by an object (as reflected by the position he chooses on an attitude scale) vary as a function of the algebraic sum of his ratings of the potency of that object for achieving or blocking the realization of his values (when the importance of these values is held constant).

3. The degree and sign of affect aroused in an individual by an object (as reflected by the position he chooses on an attitude scale) vary as a function of the algebraic sum of his ratings of the importance of the values whose attainment or blocking he perceives to be affected through the instrumental potency of that object (when the instrumental potency of that object for attaining or blocking each of these values is held constant).

METHOD

Subjects and Attitude Measure

One hundred and twenty *Ss*, recruited from undergraduate courses at the University of Michigan and the Ypsilanti State Teachers College, took an attitude questionnaire that contained, among other items, one dealing with the issue of "whether members of the Communist Party should be allowed to address the public." Each *S* checked his first choice among five alternative statements. Seventeen *Ss* chose the alternative indicating extreme opposition to allowing members of the Communist Party to address the public, 31 *Ss* the alternative indicating moderate opposition, 44 *Ss* that indicating moderate approval, and 25 *Ss* that indicating extreme approval. Only three *Ss* chose the

[1] This paper is based upon part of a doctoral dissertation submitted to the University of Michigan. The study was supported by the U. S. Air Force, Human Resources Research Institute, under contract #AF33-(038)-26646. The advice of Professor Helen Peak was of great value during all phases of the investigation.

alternative indicating "neutrality" on the issue, and, because of their small number, were excluded from the research population. In a retest of 95 *Ss* on this measure after an interval of at least two months, a reliability coefficient of .72 was obtained ($p < .001$).

Value Measures

Three to five weeks after the administration of the attitude measure two card-sorting tasks were administered individually. These tasks required *S* to categorize each of a group of value items in terms of (*a*) *value importance*, i.e., its importance to him as a "source of satisfaction" and (*b*) *perceived instrumentality*, i.e., his estimate as to whether, and to what extent, the value in question would tend to be achieved or blocked through the "policy of allowing members of the Communist Party to address the public."

TABLE 1

VALUE ITEMS USED IN CARD-SORTING TASKS*

1. People sticking to their own groups.
2. People looking out for the welfare of others.
3. Being looked up to by others.
4. Change and variety; having new kinds of experience.
5. Sticking to a difficult task; solving difficult problems.
6. Making one's own decisions.
7. People being strongly patriotic.
8. Serving the interests of the group to which one belongs.
9. Giving expression to feelings of anger or hostility.
10. Keeping promises made to others.
11. Having one's family approve of one's views.
12. The uncompromising administration of punishment to anyone who deserves it.
13. Having interesting work to do.
14. Having power and authority over people.
15. Being well-informed about current affairs.
16. Having the value of property well protected.
17. People of different backgrounds getting to know each other better.
18. All human beings having equal rights.
19. Being good-looking; having attractive face, body, or clothes.
20. Having a steady income.
21. Believing in a relationship between the individual and some higher spiritual power.
22. America having high prestige in other countries.
23. Being liked or loved by the opposite sex or associating with the opposite sex.
24. Getting advice on important problems.
25. People having the right to participate in making decisions which will affect them.
26. Achieving superiority over others in such things as knowledge, work, or sports.
27. Complying with the wishes of persons in authority.
28. Being like others in general; having the same interests.
29. People being well educated.
30. People having strict moral standards.
31. The open expression of disagreement between people.
32. Everyone being assured of a good standard of living.
33. Being allowed to maintain the privacy of one's opinions and beliefs.
34. Being with other people; socializing.
35. Letting others make their own decisions.

* This list does not include the "salient" value items specially constructed for each *S*.

The pack of value-cards included 35 items constructed in the light of White's value-analysis technique (10) and Murray's analysis of major needs (6). (The value items used are given in Table 1.) In addition to these 35 items, the pack of cards presented to each *S* also contained value terms that had been coded out of his questionnaire answer to a verbal probe in conjunction with the item on "allowing members of the Communist Party to address the public." For most *Ss* the total number of such "salient" values came to two or three, the range among all *Ss* extending from zero to six.

For the "value importance" measure *S* was asked to rank each card so as to indicate how much satisfaction he gets, or would get, from the value state that it described. Each value was to be judged independently in terms of 21 categories ranging from "gives me maximum satisfaction" (Category +10) through "gives me neither satisfaction nor dissatisfaction" (Category 0) to "gives me maximum dissatisfaction" (Category −10). For the measure of "perceived instrumentality" he was asked to judge and place each card in terms of 11 categories ranging from "the condition (value described on the card) is completely attained by allowing admitted Communists to address the public" (Category +5) through "whether or not admitted Communists are allowed to address the public is completely irrelevant to the attainment of the condition" (Category 0) to "the condition is completely blocked by allowing admitted Communists to address the public" (Category −5). In each of the two card-sorting tasks *S* was required, after the initial categorization was completed, to rank the cards *within* each category in terms of "value importance" and "perceived instrumentality" respectively.

The test-retest reliability of these two measures was studied with a subpopulation of 12 *Ss*, the second administration following the first by from four to five weeks. For each of these *Ss* Spearman's rho was computed as an estimate of the degree of correlation between the ranks assigned to the separate values on the first and second administrations respectively. For the measure of value importance, the median rho was .89 ($p < .01$); for that of perceived instrumentality, it was .74 ($p < .01$).

From the data obtained through the two card-sorting tasks ten indices were computed to permit testing the major hypotheses. Operational definitions of these indices are given in Tables 2–4.

RESULTS AND DISCUSSION[2]

Each hypothesis was tested by chi square,[3] computed from a 3 × 4 table. In each such

[2] All of the results reported in this section have been successfully replicated with regard to the attitude area of "allowing Negroes to move into white neighborhoods." The replication was based on an analysis exactly parallel to the one described in the present report and upon identical measurement operations.

[3] This technique was used rather than correlational methods because it was felt that the assumptions required by correlational methods (particularly the

table the four groups in terms of attitude position (extremely favorable, moderately favorable, moderately unfavorable, and extremely unfavorable toward "allowing members of the Communist Party to address the public") were cross classified in terms of a threefold categorization of the *S*s on one of the indices derived from the card-sorting value measures.

Hypothesis 1

All of the four indices used to test Hypothesis 1 provide, as the hypothesis requires, for representation of both the importance to the individual of some set of values and his perceptions as to how these values are affected (with regard to their attainment or blocking) through the instrumental agency of the attitude object (the policy of "allowing members of the Communist Party to address the public"). By algebraically summing the importance-instrumentality products for each of the values it is possible to represent appropriately the interaction between positive and negative values on the one hand and perceptions of positive and negative instrumentality on the other. Thus a positive value (rated say as +7 in its capacity to satisfy the *S*) may be perceived as being attained through the policy of "allowing members of the Communist Party to address the public" (to the extent say, of a rating of +3 on "instrumentality"). In this case the product would be +21. If on the other hand, the same value were rated as being *blocked* by the policy of "allowing members of the Communist Party to address the public" (for example a rating of −3 on "instrumentality") the resultant product would be −21. Similarly, a value rated as yielding dissatisfaction and thus bearing a minus sign on "value importance" would yield a negative product if multiplied by a rating of positive instrumentality and a positive product if multiplied by a rating of negative instrumentality. By summing all such products for a given set of values a single algebraic quantity is obtained representing the total import of the *S*'s pattern of beliefs about the influence of the attitude object upon attaining or blocking various states that he values to differing degrees.

TABLE 2

RELATIONSHIPS BETWEEN "VALUE IMPORTANCE × PERCEIVED INSTRUMENTALITY" (INDICES 1–4) AND ATTITUDE POSITION (N = 117)

Index	Description of Index	Chi Square	Probability*
1	Algebraic sum of importance-instrumentality products for all value items	26.33	.001
2	Algebraic sum of importance-instrumentality products for the twenty values ranking highest on importance	30.82	.001
3	Algebraic sum of importance-instrumentality products for all "salient" values	45.09	.001
4	Algebraic sum of importance-instrumentality products for all "non-salient" values	15.10	.02

* Examination of the actual data plots reveals that all of these significant chi squares are due to positive, monotonic relationships.

Since all four of the indices used to test Hypothesis 1 were found to be significantly related to attitude position (see Table 2), the hypothesis would appear to have been confirmed at an acceptable level. These findings, to generalize, lend support to the view that beliefs associated with an attitudinal affect tend to be congruent with it; i.e., that there exists within the individual an "organization" of the affective and cognitive properties of his total pattern of response to what is, for him, an "attitude object."

One possible objection to such an interpretation might hold that the way in which *S* sorts the cards may reflect not his actual cognitive structure but rather an attempt to demonstrate to the experimenter and to himself that his beliefs are consistent with his attitudes. If striving for cognitive-affective consistency accounts in part for relationships obtained during the testing session, however, it seems appropriate to assume that similar motivation had been in operation before the testing session, i.e., that the motive to "rationalize" one's attitudes plays a role in the natural history of attitude development.

Still another line of evidence that limits the significance of this objection arises from the findings obtained with Index 3. This index is based solely upon the "value importance" and "perceived instrumentality" ratings of the *S*'s

assumptions of normality and equal-interval scales) were not fully met by the data. The method used for computing chi square was the "maximum likelihood test" of Mood (5) which is applicable to data plots with low cell counts.

"salient" values, which were coded out of his own verbal defense of his stated attitude position. Certainly the individual's estimates of the importance of these values and the ways in which their realization would be influenced by the attitude object must have had *pre-experimental* reality to have been given expression in response to the questionnaire probe.

It should be noted, however, that the historical relationship between affective and cognitive processes in the development of attitudes is not approachable through analysis of the present data. The province of the present study is to examine the relationships between cognitive and affective processes at a point in time after an attitude has been acquired and stabilized. Nevertheless, the present hypotheses may, if confirmed, suggest further hypotheses about the processes of attitude learning and attitude change.

Hypotheses 2 and 3

While the data just reviewed confirm the existence of a relationship between a measure of affect toward an object and a measure based upon both the "importance" and "perceived instrumentality" of values associated with it, the question may be raised as to whether each of these two latter variables may be demonstrated to covary with affect when the other is held constant. Hypotheses 2 and 3 assert the existence of such relationships. Indices 5–10 are based upon procedures for obtaining separate estimates of the "value importance" and "perceived instrumentality" variables. To isolate the one variable, it was possible to hold the other constant since the original card-sorting tasks had required *S* to rank the values *within* each of the categories used in sorting.

Thus, to test the relationship between "perceived instrumentality" of values and affect toward the attitude object, four indices (5–8) were developed, all of which have in common being based upon a summation of the "perceived instrumentality" ratings of a group of values defined and chosen in terms of their "value importance" ranks. Index 5 for example is obtained by choosing the value items ranked 1–20 by the *S* on the dimension of "value importance" and then obtaining the algebraic sum of the "perceived instrumentality" ratings of these same 20 values.

A similar procedure was employed with regard to the indices (9 and 10) used to test the relationship between "value importance" and affect toward the attitude object. These two indices are both based upon the "value importance" ratings of a group of values defined and chosen in terms of their "perceived instrumentality" ranks. Index 9 for example is obtained by choosing the five values ranked highest on "perceived positive instrumentality" and then obtaining the algebraic sum of the "importance" ratings of these same five values. Index 10 is identical except that it is computed from the five values ranked highest on "perceived negative instrumentality."

The results reported in Table 3 indicate that all four of the indices of "perceived instrumentality" are significantly related to attitude position. Furthermore it should be noted that the hypothesis has been verified at a number of different levels of "value importance" control, each of the four indices representing a different level. On the basis of these findings Hypothesis 2 appears to have been confirmed. To generalize, other things (particularly "value importance") being equal, extreme attitudinal affects are associated with perceptions of close positive or negative instrumental connection between the attitude object and related values, while moderate attitudinal affects are associated with perceptions of less clear-cut instrumental relationships. In terms of attitude theory this finding suggests the existence of a distinguishable and important

TABLE 3

RELATIONSHIPS BETWEEN "PERCEIVED INSTRUMENTALITY" (INDICES 5–8) AND ATTITUDE POSITION (N = 117)

Index	Description of Index	Chi Square	Probability*
5	Algebraic sum of the instrumentality ratings of the values ranking 1st–20th on importance	35.51	.001
6	Algebraic sum of the instrumentality ratings of the values ranking 11th–30th on importance	21.27	.001
7	Algebraic sum of the instrumentality ratings of the values ranking 1st–10th on importance	20.82	.01
8	Algebraic sum of the instrumentality ratings of the values ranking 11th–20th on importance	38.47	.001

* Examination of the actual plots reveals that all of the significant chi squares are due to the presence of positive, monotonic relationships.

TABLE 4
RELATIONSHIPS BETWEEN "VALUE IMPORTANCE" (INDICES 9 AND 10) AND ATTITUDE POSITION (N = 117)

Index	Description of Index	Chi Square	Probability*
9	Algebraic sum of the importance ratings of the values ranking 1st–5th on positive instrumentality (attainment of values)	12.75	.05
10	Algebraic sum of the importance ratings of the values ranking 1st–5th on negative instrumentality (blocking of values)	16.46	.02

* Examination of the actual plots reveals that the significant chi square in the case of Index 9 is due to the presence of a positive monotonic relationship, while the significant chi square in the case of Index 10 is due to the presence of a negative monotonic relationship, both as predicted.

dimension of attitude-related cognitive structures. Recent research (1) based upon the present study lends confirmation to the existence of this dimension as a separate and manipulable one and indicates that attitude-change effects (in the sense of affective change) may be achieved through its manipulation.

The results reported in Table 4 indicate that both of the indices of "value importance" are significantly related to attitude position. On the basis of these findings, Hypothesis 3 appears to have been confirmed.

Although the relationships reported in Table 4 reach an acceptable level of significance, however, their significance level is lower than most of the relationships reported in Tables 2 and 3. The present data leave it an open question as to whether this is an artifact of the measurements or computations employed or whether, on the other hand, "perceived instrumentality" actually controls more variance in attitudinal affect than does "value importance." The generalization is nevertheless supported that other things (particularly "perceived instrumentality") being equal, extreme attitudinal affects are associated with values of high importance while moderate attitudinal affects are associated with values of less importance. For attitude theory, these results suggest the existence of a second distinguishable and important dimension of attitude-related cognitive structures. Further research is needed to determine whether the individual's assessment of the importance of his values is manipulable and whether, if it is, such manipulation is sufficient to produce change in affect toward an object seen as having instrumental influence upon the attainment or blocking of these values.

Further Implications for Attitude Theory

It has been suggested that attitudinal affect toward an object may be altered by the prior modification of value importance and perceived instrumentality. A related implication is that one way in which attitudes may originally develop is through the prior acquisition of beliefs about the value-attaining or value-blocking powers of particular objects, be those objects individuals, groups, political proposals, or commercial products. Such a view, even if it has not been systematically elaborated, seems to underlie the attitude-creating and attitude-changing communication techniques of many propagandists in such fields as advertising, teaching, psychological warfare, and child-rearing.

A less obvious implication of the present findings suggests at least one other type of sequential process leading to the establishment of stabilized affective-cognitive patterns. It seems plausible that much original attitudinal learning (the acquisition of a stable affective response where none existed previously) may originate in experiences of being rewarded or punished for imitation or rehearsal of expressions of affect provided by others. A case in point may be found in the data on the acquisition of the anti-Negro attitude as reported by Horowitz (4). With an affect, or its beginnings, established through reinforcement procedures, something like a need for affective-cognitive consistency may set the person to acquire socially available beliefs that "rationalize" the acquired affect. When such beliefs are unavailable in the person's communicative surround, he may invent them. In real-life situations where attitudes are being instilled, there are generally available both cognitive supports for the advocated affect and direct reinforcements for its expression.

Recent research by the investigator seems to indicate that the production of *change* in an already established attitude may be obtained without any direct attempt to modify the associated cognitive structure, but rather through direct assault (with hypnotic techniques) upon the established affect. Such assault produces in some *S*s a temporary affective reversal, and

along with this reversal, a spontaneous, self-directed modification of the related cognitive structure.

SUMMARY

An attitude questionnaire on the issue of "allowing members of the Communist Party to address the public" was administered to 117 *S*s. Three to five weeks later each *S* took a card-sorting test which required him to rate and rank each of a group of value items in terms of (*a*) the importance of the value as a source of satisfaction and (*b*) his perception as to the extent to which the value tends to be attained or blocked through the instrumental agency of the attitude object.

The data thus gathered were used in testing three hypotheses concerned with the relationships between stable affective response (attitude) toward an object and beliefs about that object. These hypotheses were derived from a general proposition to the effect that the sign and extremity of affect toward an object are functions both of whether it is perceived as facilitating or blocking the attainment of values and of whether or not the values involved are important ones. All three hypotheses were confirmed. Some of the data were interpreted as suggesting that "value importance" and "perceived instrumentality" are separate and possibly, manipulable dimensions of attitude-related cognitive structures.

Some suggestions were presented as to the implications of the present study with regard to the phenomena of attitude learning and attitude change.

REFERENCES

1. CARLSON, E. R. Attitude change and attitude structure. *J. abnorm. soc. Psychol.*, 1956, **52**, 256–261.
2. CARTWRIGHT, D. Some principles of mass persuasion. *Hum. Relat.*, 1949, **2**, 253–268.
3. Hilliard, A. E. *The forms of value*, New York: Columbia Univer. Press, 1950.
4. HOROWITZ, E. L. The development of attitude toward the Negro. *Arch. Psychol., N. Y.*, 1936, No. 194.
5. MOOD, A. E. *Introduction to the theory of statistics.* New York: McGraw-Hill, 1950.
6. MURRAY, H. A. *Explorations in personality.* New York, Oxford Univer. Press, 1938.
7. ROSENBERG, M. J. The experimental investigation of a value theory of attitude structure. Unpublished doctor's dissertation, Univer. of Michigan, 1953.
8. SMITH, M. B. Personal values as determinants of a political attitude, *J. Psychol.*, 1949, **28**, 477–486.
9. TOLMAN, E. C. A psychological model, in T. Parsons & E. A. Shils, (Eds.), *Toward a general theory of action.* Cambridge: Harvard Univer. Press, 1951.
10. WHITE, R. K. *Value analysis: the nature and use of the method.* Society for the Psychological Study of Social Issues, 1951.
11. WOODRUFF, A. D. Personal values and the direction of behavior. *Sch. Rev.*, 1942, **50**, 32–42.
12. WOODRUFF, A. D., & DIVESTA, F. J. The relationship between values, concepts, and attitudes. *Educ. psychol. Measmt.*, 1948, **8**, 645–660.

Received January 23, 1956.

THE EFFECTS OF EMOTIONAL AROUSAL ON THE RETENTION OF FILM CONTENT: A FAILURE TO REPLICATE

ELEANOR E. MACCOBY, HARRY LEVIN,[1] AND BRUCE M. SELYA

Harvard University

IN A preliminary report (2) we described a study in which groups of children were subject to experimental variation in their emotional states before being shown a movie. The nature of the movie content recalled by the children a week later was reported to be a function of their emotional state at the time of viewing. In an attempt to replicate the study, however, we have since obtained negative results. Since the original abstract has been rather widely cited,[2] it seems desirable to publish these negative findings.

The hypothesis tested in the two studies was: if children are frustrated just before they see a movie, so that aggressive impulses are aroused, they remember more of the aggressive content of the film than do children who are not frustrated. This hypothesis derives from the assumption that perception is to some degree selective, so that subjects who are aggressively aroused perceive more of, or attend more vigilantly to, the aspects of the material that are congruent with their drive state.

An alternative formulation, that derives the same hypothesis from other premises, concerns the drive-reducing function of fantasy. The assumption that wish-fulfillment fantasies are satisfying is shared by many personality theorists, and Feshbach (1) has demonstrated that at least some kinds of fantasy may reduce an experimentally aroused drive. If under conditions of frustration, aggressive content is more drive reducing than other content, the present hypothesis could be derived in terms of either of two possible mediating processes: (*a*) The aggressive actions of fantasy heroes, which presumably provide some drive reduction to the viewers, serve as stimuli that may form habits with some implicit, probably verbal responses, so that under recall conditions their recall probability is higher than for those fantasy stimuli that are not associated with drive reduction; or (*b*) Possibly the drive-reducing aspects of the movie acquire thereby a pleasant affective tone, thus increasing the chances of spontaneous practice of the material during the interval before the recall test.

The following report describes the *second* in the series of two experiments.

METHOD

Subjects

The experiment was conducted in two semirural schools in upper New York state. The subjects were the children in eight fifth- and sixth-grade classrooms, numbering 190 children in all.

Procedure

Frustration. The first experimenter, a man, was introduced to the class as an expert on spelling. He told the children that there was to be a statewide spelling contest, that he was about to run a spelling bee in their class, and if the winning team were good enough, they might have a chance to enter the contest. The class was divided into two teams, the "reds" and the "blues," and the members were given appropriately colored ribbon badges. The two teams in each classroom were matched for sex and spelling proficiency as rated by the teachers, and a random procedure was used to determine which team was red, which blue. During the spelling bee, each child was allowed two mistakes before he had to sit down. The red team was given ninth-grade words to spell, and the blue team got fourth-grade words. The experimenter was cold, critical, and sarcastic toward the red team and friendly and supportive toward the blue team. The red teams were spelled down quickly, and began to complain of unfair treatment. The experimenter asked them if they thought they could do better if they had a second chance. They were then permitted to repeat the spelling bee, with the blue team continuing to receive favored treatment. At the direction of the experimenter, a member of the winning team collected the badges of the losing team, while the winning blues were permitted to keep theirs. (The fact that the children were highly involved in the contest is suggested by the fact that one little girl on a winning team got up early every morning for a week to iron her badge so she could wear it to school!) In some classrooms, the meaning of the colors was reversed, so that the reds were favored over the blues. After the badges of the losing team had been collected, the first experimenter left the room.

Exposure to the film. Toward the end of the spelling bee the second experimenter (a woman) came into the room and sat at the back of the room. After the spelling

[1] Now at Cornell University.

[2] See for example the report of the hearings of the Kefauver Committee on Juvenile Delinquency, specifically the hearings on the effect of television programs, April 6 and 7, 1955.

bee, she came forward and apologized for coming on such a busy day. (Every effort was made to dissociate the spelling bee and the movie.) She told the children that she was interested in finding out what sorts of things children liked and remembered in movies, and said she had a movie for them to look at, mentioning that she might ask them a few questions about it afterwards. The first chapter of a serial movie was then presented, starring the Dead End Kids and the Junior G Men. At the end of the movie which involved considerable aggressive action, *E*-2 told the children that there was no time for discussion now, and promised to come back in a week with another chapter of the film.

Measures for memory of movie content. A week later, the children were given a 66-item multiple-choice recognition test on the content of the movie they had seen. Items were of the following four kinds of items: (*a*) Seventeen *aggressive central* items, pertaining to the occurrence of overt aggressive acts, the identity of the aggressor, or the identity of the recipient of aggression. (Example: "In the fight between Billy's gang and the Junior G Men, who struck the first blow?" (*b*) Twelve *aggressive incidental* items, concerned with details surrounding aggressive acts. (Example: "In the fight that takes place immediately after the two trucks smash, where do the people watching the fight stand?") (*c*) Eighteen *neutral central* items pertaining to content that was central to the theme of the story, but did not involve aggressive action. (Example: "Why was Billy's father important?," the answer being that he had invented a new kind of explosive.) (*d*) Seventeen *neutral incidental* items concerned with content related to nonaggressive material incidental to the main story theme. (Example: "Jim Bradford had a very pretty secretary in his office. What was her name?")

Following the test, which took about 40 minutes, the children were shown the final chapter of the serial and were given an explanation of the experiment.

RESULTS

There were no significant differences between the frustrated and nonfrustrated groups in their recall of any of the different categories of film content. The team, rather than the individual child, was taken as the unit of analysis. Contrary to the hypothesis, there was a slight tendency for the nonfrustrated teams to remember more aggressive central content, but while six classrooms showed a difference in this direction, two showed a difference in the opposite direction. Neither a sign test, a *t* test, nor analysis of variance revealed a significant contribution of the experimental treatment to the children's recall of central aggressive content. Similarly, the frustrated and nonfrustrated subjects did not differ in their total recall scores, nor in the recall of aggressive incidental, neutral central, or neutral incidental content.

COMPARISON OF THE TWO EXPERIMENTS

The procedure for the two experiments was as nearly identical as possible. The identity of the two experimenters was different, and in the first experiment both experimenters were men. Our impression is that in both experiments, the experimental variable "took," in the sense that the children seemed much involved in the spelling bee and were anxious to win.

In the first experiment, the subjects were children living in a suburban section of the Boston area, while in the second, they were rural and small-town children in New York state. It is possible that the latter group have been less permissively treated with respect to aggression. And possibly, selective recall is a function not only of the amount of aggressive motivation present at a given time, but also of the child's level of aggression anxiety (growing out of past punishment for aggression). Presumably, if aggression anxiety is great enough, it leads to repression of aggression-related content, at least among some children. If the children in the second experiment had been more severely socialized with respect to aggression, then, this fact might account for their failure to remember more aggressive content under conditions of aggression arousal. We cannot test this hypothesis directly, since we have no information concerning the severity of socialization of aggression in these two groups of children. An indirect test is available, however. If the frustrated children in our second experiment were subject to a fairly high level of aggression anxiety, these children should be more *variable* in their recall performances than their nonfrustrated counterparts. Examination of the data shows that this was not the case. Therefore, we have no evidence for the hypothesis that the two sets of subjects differed in their level of aggression anxiety, although the possibility still exists.

REFERENCES

1. FESHBACH, S. The drive-reducing function of fantasy behavior. *J. abnorm. soc. Psychol.*, 1955, **50**, 3–12.
2. MACCOBY, ELEANOR E., LEVIN, H., & SELYA, B. M. The effect of emotional arousal on the retention of aggressive and nonaggressive movie content. *Amer. Psychologist*, 1955, **10**, 359. (Abstract)

Received June 25, 1956. Prior publication.

CASE REPORT

THE RESPONSE OF SCHIZOPHRENIC PATIENTS TO A TELEVISED WORLD SERIES GAME: A STUDY IN SOCIAL ISOLATION

JEROME L. SINGER AND HERBERT E. SPOHN[1]

Franklin D. Roosevelt V. A. Hospital, Montrose, N. Y.

PERHAPS the most typical of all the manifestations of a chronic schizophrenic state is the social isolation, the withdrawal from even the most rudimentary aspects of common social interchange. Individual interviews with patients from a schizophrenic ward may yield considerable heterogeneity in the appearance of psychotic symptomatology and personality characteristics, ranging from florid hallucinatory or delusional trends through monosyllabic response or silliness. Observation of these same patients as a social group in a ward dayroom or in various group settings suggests considerably more homogeneity, however, in the relative absence of what might be considered normal gregariousness or group behavior. Even in hospital settings (such as the one in which this study was carried out), where the personnel-patient ratio is not too unfavorable and where a full schedule of therapeutic and recreational activities is in effect, the absence of social intercourse, of mutually shared activities of a spontaneous sort, is striking. A ward dayroom of forty patients viewed on a typical afternoon following return from activities might reveal perhaps seven or eight men in conversation or engaging in a game of cards, while the rest of the group is scattered around the room with practically no sign of social interchange. Some patients may be moving restlessly, others talking in autistic fashion, others reading, others in catatonic posture, and some sleeping. One need only think by contrast of the hubbub that characterizes the dayroom of an army barracks or the courtyard of a prison to sense the dramatic depth of the social isolation of the schizophrenic. The brief study to be described here represents an initial effort to demonstrate this social isolation in a fairly objective manner and to test certain hypotheses relating social isolation to chronicity and severity of illness. As part of a long-term project dealing with the social interaction patterns of chronic hospitalized psychotics, it represents one attempt to evaluate the response-potential of schizophrenics in various group situations.

The use of reaction to a televised World Series game was dictated by the well-known response of normal males in the same age and cultural groupings as the patients to such a stimulus. Whether the audience is in the ball park or in front of a television screen, the excitement engendered by a Yankee-Dodger series quickly lowers barriers to communication among strangers. One can delineate responses to the game, e.g., cheering, hooting, or crying encouragement at players, etc., or responses to others in the crowd, betting, praising, or damning the others' favorite, or arguing the merits of a player. The questions posed in this study were: (*a*) Would the normal enhancement of interaction and social responsiveness characteristic of World Series viewing be observable in schizophrenics? and (*b*) Would it be possible to demonstrate a progression in interaction from the most severely disturbed chronic wards through the less chronic and open wards, the latter group showing a response more like that of a non-hospitalized population?

PROCEDURE

Employing a standardized rating sheet and guide, eight clinical psychologists who were stationed on eight previously-selected wards of a large mental hospital observed the response of patients during three innings of the televised third game of the New York Yankees-Brooklyn Dodgers World Series in 1955. Since the Yankees had taken a two-game lead in the series, this game might be expected to be most important. The eight wards selected were Ward A, a chronic neurological ward in which most of the patients were not psychotic and hence might serve to some extent as a control group of chronic hospitalized individuals; Ward B, the open ward of the continued treatment service, the last stop before discharge, as a rule; Ward C, the acute

[1] The authors wish to express their indebtedness to Mrs. Mildred Rosenberg, Drs. Harold Wilensky, Frank Wexler, Ralph Colvin, Frank Lachmann, and to Mr. Frank Kennedy, who served as observers for the study.

TABLE 1

PERCENTAGES OF PATIENTS AND PERSONNEL WATCHING AND INTERACTING DURING TELEVISED GAME

Ward	Percentage Watching		Percentage Interacting		Tension Level	
	Patients	Personnel	Patients	Personnel	Patients	Personnel
A	90	100	49	100	Moderate	Moderate
B	90	100	47	100	High	Moderate
C	59	100	10	100	Moderate	Moderate
D	27	100	20	80	Moderate	High
E	34	100	0	100	Low	Moderate
F	31	100	20	100	Low	High
G	24	100	0	67	Low	Moderate

service ward for new admissions, a group generally younger than the hospital mean and characterized by recent onset and relatively florid symptomatology; Ward D presumably the most advanced of the continued treatment service wards with a greater number of patients participating in hospital industrial therapy activities; Wards E and F two wards of chronic patients of approximately equal severity of over-all pathology; Ward G, a disturbed ward of chronic patients who were occasionally assaultive or elopement threats; and Ward H, a ward of elderly chronic schizophrenic patients, many of them incontinent. It was expected that the response to the game in terms of numbers watching, overt reaction in terms of cheering, etc., and interaction among viewers would decline systematically from Ward A to H. The rating sheet called for the observer to record data during the first, fifth, and ninth innings on the number of patients and staff present (the TV sets were located in ward dayrooms), the number actually *watching* the game (rather than occasionally staring up at the set), the seating arrangement, the amount of interaction between patients and between personnel, the appropriateness of response (cheering during exciting plays as against cheering during a lull for a team not in the game), the rate of tension and excitement during suspenseful moments among patients and hospital personnel, and the various modes of interaction among the viewers. From these observations it was possible to calculate data on the ratio of actual watchers to potential audience and the amount of social interaction within the audience. It should be noted that each ward contained approximately forty patients and that means for age and chronicity of illness were approximately the same for all wards except Ward C. The neurological control group of Ward A were slightly older but with similar duration of hospitalization as Wards B, D, E, F, and G.

RESULTS

It proved possible to obtain adequate ratings for seven of the eight wards in question. On Ward H the aides reported that since no interest in the game could be elicited, they had tuned the set to a cowboy movie, which was not, however, eliciting any apparent response from the patients at the time of observation. Results for the other seven wards are presented in Table 1. Data on spontaneous seating arrangements were not available since on all wards the aides arranged most of the chairs in rows in front of the TV sets while the patients were at lunch and, hence, more or less enforced a seating arrangement.

It is apparent from consideration of Table 1 that on the whole the game drew an unusually small audience. The percentage of those watching falls roughly into three groups, the neurological and open-ward patients showing a high viewing percentage, the young, acute schizophrenics, a moderate response, and the chronic schizophrenics generally undifferentiated in showing a very low response. Differences within a ward for the three innings were so slight as to warrant combination of the inning data. When the number of patients interacting with each other or the game was considered, the results are even more striking in demonstrating the absence of social response. The largest percentage of overt response emerged again from Wards A and B, but even here the percentages were low by comparison with the nearly 100 per cent overt response from the small number of attendants and other personnel who were watching the game in the dayrooms. Excitement as manifested verbally and motorically varied in similar fashion with relatively little evidence of reaction from the patients at moments which brought loud cheering from the ball-park audience and the attendants in the dayroom. No striking differences in appropriateness of response as a group occurred between the patient or attendant audiences, although on occasion a purely autistic reaction did come from a patient.

A number of qualitative observations may help to fill out the picture of severe isolation suggested by these results. Since the chairs had been oriented toward the TV set by the aides, the patients on returning from lunch automatically grouped themselves in relationship to the chairs. Thus, an initial glance into a ward suggested a much larger audience than turned out to be the case, since many patients who ordinarily stood next to chairs against the wall facing into the dayroom now oriented themselves toward the TV set. Closer observation revealed that only a few patients seated in front were actually watching the game. This impression was confirmed by inter-

viewing patients individually later in the game. Occasionally the picture and sound, perhaps because of their figural quality in the general configuration of the room, evoked an upward glance from patients who otherwise showed not the slightest awareness of the nature or course of the game. It should be stressed that many of these patients were able to verbalize appropriately on other topics in response to inquiry. Among the chronic patients who responded to the game on Wards D, E, F, and G, a high percentage turned out to be persons with borderline diagnoses of schizophrenia; they tended to be persons with "character disorder" or "alcoholism" as part of their classification and were less clearly schizophrenic in their other symptomatology.

A dramatic example of the contrast between patients and hospital personnel was the situation on Ward D. There were three barbers at work in the dayroom cutting hair during the game. They kept up a running patter of wisecracks, comments about various players' skill, and wagers about the outcome of the game. By contrast, of the dozen or so patients watching the game, only two spoke at any time or cheered appropriately during exciting moments. These two patients occasionally made comments to the others and received no response. The barbers on occasion exchanged comments with these two patients about the game, but when in their enthusiasm they made remarks to other patients or slapped them on the back, they elicited no response. When Campanella hit a home run, the barbers and the two patients shouted loudly, but there was no other cheering in the room. Apparently in response to the tension and shouting, certain patients who were obviously not looking at the TV set muttered to themselves and moved about restlessly or went through manneristic rituals. Some type of social communication passed through the patients on the ward, but at a decidedly rudimentary level of complexity. An example of this primitive social awareness was the way in which one patient, as another rose from the barber chair, walked silently from a separate part of the large room to take his seat for a haircut.

DISCUSSION

The problem of understanding how human beings who are very much alive in terms of retention of intellectual functions and, in individual interviews, sensitivity to misunderstanding or rejection, can so completely withdraw their interest from others in social groups is highlighted by these findings. Lack of interest in sports was never a striking feature of the premorbid histories of these men as far as could be determined, and so it cannot account for the apathy and absence of interaction. Nor can it be asserted that these patients have lost themselves in daydreaming or private worlds of conscious reverie because interviews and psychological tests of fantasy-disposition suggest that chronic schizophrenic patients engage in relatively little internal meditation. Perhaps chronic anxiety and fear of attack or rejection, as well as the unsatisfying nature of fantasy, have led to a kind of freezing of all modes of expression. Possibly, if one feels that some energy outlet is essential, it may take the form of endless concentration on trivial or commonplace stimuli, e.g., watching the ash burn down on a cigarette tip, noting the gurglings of one's own digestive system, or studying a repetitive drapery design.

Whatever the manner in which the chronic schizophrenic manages to maintain his isolation, the *fact* of his immense alienation from social interchange is highlighted by this study. Attempts at establishment of a therapeutic milieu through recreational and related facilities in the treatment of large numbers of patients are obviously essential in maintaining a "normal" atmosphere for patients. Yet they may be expected to prove frustrating until such time as social or clinical research provides some comprehension of the manner in which human beings can so fully detach themselves from interpersonal communication in group situations.

Received February 1, 1956.

CRITIQUE AND NOTES

AUTHORITARIANISM AND INTERPERSONAL PERCEPTION[1]

W. H. CROCKETT AND THOMAS MEIDINGER

Kansas State College

RECENTLY Scodel and Mussen (7) and Scodel and Freedman (6) have examined the relationship between authoritarianism, as measured by the California F scale, and accuracy of interpersonal perception. Their results indicate that individuals who make high scores on the F scale tend to estimate that their peers will also score high on the F scale, whether these peers are actually high or low. On the other hand, individuals who make low scores on the F scale are less uniform in their judgments, but tend to estimate peers as being either moderate or high in their responses to the F scale. No differences were found in the *accuracy* of estimates by high F and low F subjects. The present paper reports a replication of these studies, using a different population and a different operational definition of accuracy of judgment.

METHOD

Subjects (*Ss*) were drawn from three classes in general psychology at Kansas State College, and from the 80 residents of the men's dormitory. Students in general psychology were required to participate as part of their course work, while dormitory men were asked to volunteer. Of the 80 dormitory men asked, only one refused because of a lack of time.

The experimental design paralleled that used at Ohio State University by Scodel and Mussen and by Scodel and Freedman. Pairs of *Ss* were asked to discuss for 20 minutes a topic suggested by the experimenters. Each *S* had previously responded to the F scale. After the discussion *Ss* were asked to fill out the F scale as they felt their partners in the discussion would respond to it.

The *Ss* were assigned to pairs according to their free hours, No *Ss* were paired who lived at the same address or in the same section of the dormitory. Members of minority groups were excluded from the experiment. Ninety pairs of *Ss* were used. In 30 pairs the *F*-scale scores of both members were above the median of their sex; in 30 other pairs the scores of both members were below the median of their sex; and in the remaining 30 pairs one *S* scored above the median of his sex while the other scored below.

Although this study was patterned after the Ohio State studies, the following modifications in experimental design and analysis were introduced:

a. Measurement of accuracy. On the Ohio State studies accuracy is operationally defined as the extent to which the *S* estimates his partner's *total score* on the F scale. Thus, if the partner's total F-scale score is 127, the *S* is termed "accurate" if his estimate of the partner is very close to 127 and "inaccurate" if his estimate deviates markedly from 127. It has recently been pointed out (3, 5) that such a definition of accuracy has severe limitations, since any given total score on a multi-item test (excepting the highest and lowest possible scores) may be achieved in many possible ways. Since there are empirical as well as theoretical grounds for the existence of several factors in the F scale (1, 2), it seems likely that personality differences may be reflected in the *pattern* of an *Ss* responses to the individual items. An alternative definition of accuracy in interpersonal perception therefore seems desirable: the degree to which the *S* reproduces the pattern of his partner's responses. Accordingly, the profile-similarity test presented by Cronbach and Gleser (4) was used in this study as an additional measure of accuracy.

b. Questionnaire used and scoring method. The Ohio State studies employed a 28-item version of the F scale, while the present study utilized the 30-item scale presented by Adorno *et al.* (1). The method of scoring used in this study also differed from that used in the Ohio State studies. The F-scale scores in the Ohio State studies were obtained by adding a constant of 4 to the value which the *S* had assigned to each F item. Thus, a -1 response was scored 3, and responses of $+1$, $+2$, and $+3$ were scored 5, 6, and 7 respectively. In the present study, a -1 response was scored 3, as in the Ohio State studies, but responses of $+1$, $+2$, and $+3$ were scored 4, 5, and 6 respectively.

These differences in the items used and in the scoring of items probably would not alter the relative positions of respondents in their respective groups. However, the differences do rule out direct comparisons of the scores of *Ss* at Ohio State with those at Kansas State.

c. Discussion topic. In the Ohio State experiments all pairs of *Ss* were assigned "radio, television, and movies" as a discussion topic. The *Ss* in the present experiment were divided into two groups, one of which discussed radio, television, and movies while the other discussed child-training practices. Since the topic of discussion had no effect upon accuracy of judgment,[2] results from these groups will be combined in the present report.

d. Range of respondents included. All *Ss* in the Ohio State studies made either extremely high or extremely low scores on the F scale. The present experiment included as *Ss* moderate as well as extreme respondents. However, to increase comparability, we present only results based upon estimates by those *Ss* who scored in the highest and lowest quartiles.

[1] Portions of this paper were presented at the September, 1955, meetings of the American Psychological Association. Financial assistance in conducting this study was provided by the Faculty Research Fund of the School of Arts and Sciences, Kansas State College.

[2] This aspect of the experiment will be reported fully in a separate publication.

TABLE 1

ACTUAL AND ESTIMATED F-SCALE SCORES IN THE SCODEL-MUSSEN STUDY, THE SCODEL-FREEDMAN STUDY, AND THE PRESENT STUDY

Group	Actual Score of Judging Group		Estimate of Partner's Score		Diff.	t	p
	Mean	*SD*	Mean	*SD*			
High F Judging High F							
Scodel-Freedman (N = 24)	125.46	7.68	122.33	12.13	3.13	1.07	.20
Crockett-Meidinger (N = 29)	125.52	5.58	122.38	9.78	3.14	3.14	.01
High F Judging Low F							
Scodel-Mussen (N = 27)	129.26	9.34	124.19	11.08	5.07	1.79	.10
Crockett-Meidinger (N = 16)	123.25	5.04	123.81	10.01	.56	.46	.30
Low F Judging Low F							
Scodel-Freedman (N = 24)	67.04	11.94	99.75	18.06	32.71	7.42	.001
Crockett-Meidinger (N = 30)	83.13	10.71	99.10	29.56	15.97	6.46	.001
Low F Judging High F							
Scodel-Mussen (N = 27)	66.52	11.15	99.81	21.88	33.29	7.05	.001
Crockett-Meidinger (N = 14)	87.50	8.29	101.14	9.95	13.64	6.85	.001

TABLE 2

FREQUENCY OF ESTIMATED F-SCALE SCORES ACCORDING TO COMBINATION OF ACTUAL SCORES IN THE SCODEL-MUSSEN STUDY, THE SCODEL-FREEDMAN STUDY, AND THE PRESENT STUDY

Group	Estimated F-Scale Score		
	High*	Middle*	Low*
High Judging High			
Scodel-Freedman	16	8	0
Crockett-Meidinger	20	9	0
High Judging Low			
Scodel-Mussen	26	1	0
Crockett-Meidinger	12	4	0
Low Judging High			
Scodel-Mussen	9	13	5
Crockett-Meidinger	1	7	6
Low Judging Low			
Scodel-Freedman	6	14	4
Crockett-Meidinger	4	15	11

* In the Scodel-Mussen study *Ss*' estimates were called high if the total score was 119 or more, medium if the score was between 80 and 118, and low if it was 79 or below. In the Scodel-Freedman study the figures were high, 115 and over; medium, 80–114; and low, 79 and below. For the present study the figures for high, medium, and low categories respectively were: men, 117 and over, 97–116, and 96 and below; for women 116 and over, 96–115, and 95 and below. It should be pointed out again that the scoring technique used in the Ohio State studies was different from that used in the present study, and that the total scores are not directly comparable.

e. Sex differences in F-scale scores. Our experimental design called for one-third of the pairs in each condition to be women. The median F-scale score for men was five points higher than the median for women. If, as it seemed plausible to assume, this sex difference reflects a social norm for women considerably less "authoritarian" than that for men, assigning *Ss* to high F or low F status on the basis of the combined distribution would result in the misplacement of some *Ss* relative to the norm for their sex. Male and female *Ss* were therefore assigned to high F or low F status relative to the median of their own sex. Similarly, the distributions of profile-similarity scores were kept separately for men and women.

RESULTS

Table 1 presents for both of the Ohio State studies and for the present study the actual mean F-scale scores of high F and low F *Ss* and the mean scores that these *Ss* attributed to their partners.[8] In all of these studies, the mean estimate by high F *Ss* was very similar to the mean of the high F *Ss*' own scores, regardless of the partners' actual positions on the F scale. In one comparison in the Kansas State study—that in which high F *Ss* judged high F partners—the mean of the estimates was significantly lower than the mean of the *Ss*' own scores. Nevertheless, the absolute magnitude of the difference between these means was quite small, especially when compared with the differences between estimated and actual means of low F *Ss*. For low F *Ss*, on the other hand, the mean of the estimates of all groups was significantly higher than the mean of the low F *Ss*' own scores, regardless of the partners' actual positions.

Table 2 further clarifies the divergent estimation behavior of high and low F *Ss*. This table presents the number of *Ss* in each group whose estimates of their partners fall within the high, middle, and low categories of the F-scale distribution. As may be seen, most high F *Ss* estimated their partners to be in the highest group, and no high F *S* attributed a low F-scale score to his partner. For low F *Ss*, on the other hand, much more variability is present

[8] We are greatly indebted to Professor Alvin Scodel for making these data available for the Scodel-Freedman experiment.

in the low F estimates than in estimates by high F *Ss*.

The Ohio State experimenters have advanced as an explanation of this difference in variability between highs and lows a tendency for the authoritarians to project their own responses upon their partners. As Scodel and Mussen say, "The highs' strong need to identify with an ingroup of superior status results in perceptual distortions which exaggerate and extend the similarities between themselves and their peers" (7, p. 184).

To test this hypothesis we defined projection operationally as the similarity between the pattern of the *S*'s own responses and the pattern of his estimates of his partner. The assumed similarity scores obtained were ordered in a frequency distribution for each sex. These distributions were divided at the median, and the number of high F and low F *Ss* falling above and below the median assumed similarity score was ascertained. High F *Ss* were no more likely than low F *Ss* to have high assumed similarity scores (chi square = 1.12, $p > .20$). Thus, the projection hypothesis was not confirmed.

As an additional test of accuracy of perception, for each *S* a profile-similarity score was computed comparing his estimate of his partner with the partner's actual responses. The distributions of these profile-similarity scores for each sex were dichotomized. High F and low F *Ss* did not differ significantly in the accuracy of their estimates as measured by this method (chi square = .12, $p > .70$).

DISCUSSION

It is evident that the results of the present study are strikingly similar to those at Ohio State. A number of explanations of the results are available. One possible explanation assumes that agreement with a number of F-scale items is characteristic of many fundamentalist and conservative social, religious, and political groups in our society. A high F *S* whose social experiences have been principally within such groups and who has had little experience with vociferous proponents of contradictory views is likely to believe that a particular person will agree with him unless he has direct information to the contrary. Twenty-minute discussions probably do not provide much direct information about the partner's position on these items. On the other hand, in view of the wide dissemination given conservative, traditional views by our institutions of mass communication and by certain extremely vocal political groups, it would be unlikely for a low F *S* to avoid recognizing that a considerable number of people disagree with him about certain items of the F scale. Since the low F *S*'s attitudes have not developed in a social vacuum, however he is also likely to be familiar with individuals and groups who share his own opinions. Thus, the overestimation of the degree of authoritarianism and the greater variability of low F *Ss* may simply reflect contemporary social conditions.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. CHRISTIE, R., & GARCIA, J. Subcultural variation in authoritarian personality. *J. abnorm. soc. Psychol.*, 1951, **46**, 457–469.
3. CRONBACH, L. J. Processes affecting scores on understanding others and assumed similarity. *Psychol. Bull.*, 1955, **52**, 177–193.
4. CRONBACH, L. J., & GLESER, GOLDINE C. Assessing similarity between profiles. *Psychol. Bull.*, 1953, **50**, 456–473.
5. GAGE, H. L., & CRONBACH, L. J. Conceptual and methodological problems in interpersonal perception. *Psychol. Rev.*, 1955, **62**, 411–422.
6. SCODEL, A., & FREEDMAN, MARIA L. Additional observations on the social perceptions of authoritarians and nonauthoritarians. *J. abnorm. soc. Psychol.*, 1956, **52**, 92–95.
7. SCODEL, A., & MUSSEN, P. Social perceptions of authoritarians and nonauthoritarians. *J. abnorm. soc. Psychol.*, 1953, **48**, 181–184.

Received December 7, 1955.

ETHNIC CLEAVAGE AMONG YOUNG CHILDREN

WALLACE E. LAMBERT AND YOSH TAGUCHI
McGill University

AN APPARENT discrepancy exists between the age when preferential or prejudiced *attitudes* toward ethnically different people develop and the age when preferential or prejudiced *behavior* appears. A recent summary of research (6) places the appearance of prejudiced attitudes at the third or fourth years compared with the onset of prejudiced behavior at the eighth to tenth years. Moreno (11) found no sociometric evidence of behavioral cleavage among young children of various ethnic backgrounds until age ten (Grade 5). Criswell (2), using the same technique, studied children from nursery school to the teen ages and found no cleavage until the eighth year (Grade

3). She noted that ethnic cleavage was most pronounced at the tenth year (Grade 5). The age determined by Moreno and Criswell is not only accepted as the established age of onset of prejudiced behavior in children, but their findings are also used as supporting evidence for theories of prejudice (e.g., 6, p. 1037; 8, p. 514).

The appropriateness of Moreno's and Criswell's method is open to question. Standard sociometric questions are often not meaningful to preschool *Ss*. In preliminary work we found that four- and five-year-olds generally "liked everyone" and "didn't mind" who sat next to them, suggesting that young children may not discriminate favored associates on such issues. Discriminations may be made, however, if the more significant values of preschool children are incorporated in the choosing procedures. To test this possibility, we measured ethnic cleavage among young children with techniques which, we assume, are more meaningful to them.

METHOD

The thirteen *Ss* of this study comprised the total membership of a small nursery school in Montreal. Six children were of Japanese origin (Oriental) and seven were of Roumanian, Greek, German, and Polish ancestries (Occidental), but all thirteen spoke English fluently. They attended school from 9:00 A.M. to 5:00 P.M. five days a week since both parents of all the children normally worked in Montreal. The newest member had been in attendance for over two months when the investigation was begun. None of the children in the school lived close enough to one another to play together outside of school. Two of the Oriental and three of the Occidental *Ss* were male. The ages for Oriental *Ss* were: 5.1, 5.6, 5.6, 5.8, 5.8, and 6.0; for the Occidental: 3.5, 4.0, 4.2, 5.5, 6.0, 6.0, and 6.2.

The experimenter (Taguchi, an Oriental) was introduced to the children as a guest supervisor (one regular supervisor was a white Canadian, the other an Oriental) and actively participated in the school program once each week for a month before the investigation commenced. All experimental work with the children began at 11:00 A.M. during the recreational period.

Type of choices. Four-year-olds generally place high value on receiving or giving candy and on having their pictures taken. It was assumed that choices of associates would be meaningful to children if these values were made part of the choosing procedure. The *Ss* were asked (*a*) to give away a piece of candy to an associate and then receive one for themselves (candy-giving choices); (*b*) to choose an associate to pose with them for picture taking (picture-taking choices); and (*c*) to indicate which picture, from among individual poses of each school member, they liked best and would want to keep, excluding their own (picture-choosing choices). In each of these tests the child was taken individually to a side room and given directions for each situation. The candy-giving test was repeated twelve times during a four-week period while the other two tests occurred once only.

TABLE 1

TEST CHOICES OF OCCIDENTAL AND ORIENTAL CHILDREN

	Candy-Giving Choices*		Picture-Taking Choices		Picture-Choosing Choices	
	In group	Out group	In group	Out group	In group	Out group
Occidentals	4	3	5	2	4	3
Orientals	6	0	6	0	6	0
Exact tests of significance		$p = .048$†		$p = .002$†		$p = .048$†

* Mean choices for twelve repetitions of the procedure.
† Two-tailed tests of significance.

RESULTS AND DISCUSSION

The choices made in the three test situations are presented in Table 1. Exact tests of significance (7, p. 303f) indicate a significant association between ethnic background and type of choices made, for all three situations. However, the tendency for Occidental children to choose within their own ethnic group is not a significant one since the probabilities of obtaining differences equal to or greater than a 4–3 and 5–2 split by chance alone are, respectively, 1.00 and .453 using an exact two-tailed test. Ethnic cleavage is clearly apparent, however, with the Oriental children. The Orientals' within-group preference is significant since the probabilities of obtaining a difference equal to or greater than a 6–0 split is .031 for each instance.

The findings indicate that ethnic cleavage appears among preschool children. We maintain that other researchers have not found prejudiced behavior at earlier ages because their methods of observation have not penetrated to the significant values of the younger *Ss*. Further research is clearly necessary in order to generalize this finding beyond Occidental-Oriental *Ss* and beyond so small and compact a group as that used here.

Davitz (3) has recently suggested that people have a basic need to be similar to valued persons, this need manifesting itself in a child's striving to be like his parents or an important peer. Davitz turns to Mowrer's theory of developmental identification (12) to account for such a need. For Mowrer, identification is equated with the secondary reward value of a person for the infant. Developmental identification starts with the interaction between the infant and the others in his social environment who satisfy his needs. When drive reduction and the cues associated with those who reduce the drives are repeatedly paired, the cues themselves take on secondary reward value for the infant. On a similarity gradient, school-

mates of the same ethnic background would be more alike, at least in physical characteristics, the family members with whom the child has identified. Any secondary reward generated during the identification process would be more likely to generalize to like schoolmates.

It is apparent that Mowrer's theory accounts for the behavior of our Oriental *S*s but does not explain the Occidental group's behavior. Our finding of "one-way" cleavage is not unique. Many researchers have noted the preferential choice of within-group members with school-age children. Criswell (2) noted that, when racial cleavage did appear, it was the Negro *S*s who were the first to segregate themselves. Radke *et al.* (13) found that Jewish children, 5–9 years old, were more emotionally identified with their own group than were Catholic or Protestant children. At the high-school age, Loomis (9), Goodnow and Taguiri (5), and Lundberg (10) have all noted ethnic and religious within-group segregation.

It appears that minority group members (members of groups toward whom prejudice is commonly directed) are generally the first to segregate themselves from the majority group, this self-segregation occurring with preschool and young children as well as at the high-school level. Mowrer's and Davitz's theories can be extended to account for this empirical fact. Nurturance needs arising in infancy are satisfied by people who have distinguishable cue properties. The cues, acting alone as secondary reward agents, may become more active agents whenever nurturance-type needs are activated. Minority status, with its aspects of threat, would regenerate similar needs and concomitantly intensify the significance and vividness of cues associated with own-group members. The secondary reward accompanying the cue properties of within-group associates will reduce the nurturance needs of minority group members in a mixed-group situation. Majority group members, with less environmental threats, will not experience increased nurturance-type needs in a mixed-group situation and will accordingly receive less reward from within-group situations.

G. Allport has discussed the self-segregation phenomenon in other terms. He believes that within-group cohesion mitigates the problems of minority group members by assuring them friendship (1, p. 149), and permits them easier interaction where they need not learn new habits of social intercourse (1, p. 18).

A recent study by Goodman (4), however, poses problems for both Mowrer and Allport, and suggests that further attention should be given to prejudiced behavior among young children. Goodman found that Negro preschool children often showed preference for *white* characters in stories, white dolls, and even white people.

SUMMARY

Ethnic cleavage among preschool children was studied using various test situations where choices of associates could be observed. It was found that ethnic cleavage does exist when choices are made significant to young children. The results are compared with other related findings and attempts are made to explain them theoretically.

REFERENCES

1. ALLPORT, G. W. *The nature of prejudice.* Cambridge, Mass.: Addison-Wesley, 1954.
2. CRISWELL, JOAN H. Racial cleavage in Negro-white groups. *Sociometry*, 1937, **1**, 85–89.
3. DAVITZ, J. R. Social perception and sociometric choice of children. *J. abnorm. soc. Psychol.*, 1955, **50**, 173–176.
4. GOODMAN, MARY ELLEN. *Race awareness in young children.* Cambridge, Mass.: Addison-Wesley, 1952.
5. GOODNOW, R. E., & TAGIURI, R. Religious ethnocentrism and its recognition among adolescent boys. *J. abnorm. soc. Psychol.*, 1952, **47**, 316–320.
6. HARDING, J., KUTNER, B., PROSHANSKY, H., & CHEIN, I. Prejudice and ethnic relations. In G. Lindzey (Ed.), *Handbook of social psychology.* Cambridge, Mass.: Addison-Wesley, 1954. Ch. 27.
7. KENDALL, M. G. *The advanced theory of statistics.* London: Griffin, 1943.
8. KLINEBERG, O. *Social psychology.* New York: Holt, 1954.
9. LOOMIS, C. P. Ethnic cleavages in the southwest as reflected in two high schools. *Sociometry*, 1943, **6**, 7–26.
10. LUNDBERG, A., & DICKSON, L. Selective association among ethnic groups in a high school population. *Am. sociol. Rev.*, 1952, **17**, 23–34.
11. MORENO, J. L. *Who shall survive?* Washington: Nervous & Mental Disease Pub. Co., 1934.
12. MOWRER, O. H. *Learning theory and personality dynamics.* New York: Ronald, 1950.
13. RADKE, M., TRAGER, H. G., & DAVIS, H. Social perceptions and attitudes of children. *Genet. psychol. Monogr.*, 1949, **40**, 327–347.

Received December 1, 1955.

DIFFERENCES IN THE PERCEPTION OF AN AUTHORITY FIGURE AND A NONAUTHORITY FIGURE BY NAVY RECRUITS

WARREN J. WITTREICH[1] AND KEITH B. RADCLIFFE, JR.

Naval Medical Research Institute[2]

THE Navy recruit is introduced into military discipline immediately upon his arrival in boot camp. He is taught that orders given are orders obeyed—the alternative being severe disciplinary action. He rapidly learns to recognize those individuals who represent authority, and he becomes aware of the implications the these authority figures have for his own successful behavior.

In the early phases of training, when the recruit is still relatively unsure of himself, the appropriate visual configuration bearing stripes or bars is often sufficient to provoke fear and anxiety. As recruit training teaches him the appropriate forms of behavior in the appropriate situations, the element of fear is gradually diminished. Yet, seldom is that element completely eliminated. The comforting reliability of being in the presence of one's own peers is invariably preferred to the ever-present uncertainty involved in dealing with authority.

The question then arises: Does this element of fear involved in viewing and being with an authority figure actually influence the perception of that figure? Previous studies in this general area would indicate an affirmative answer. Gilder *et al.* (1) provide evidence which seems to indicate that a threatening figure changes less readily than a nonthreatening figure under conditions of induced optical change. Wittreich and Radcliffe (2) have shown that the human figure under the condition of simulated mutilation is less susceptible to induced optical change than the human figure under conditions of normal configuration.

Hence the hypothesis is advanced that under conditions of induced aniseikonic distortion a figure that represents authority is more resistant to perceptual change than a figure that does not represent authority. This hypothesis was tested in a study of Navy recruits.

METHOD

Twenty-four white male Navy recruits in their seventh week of training at the Naval Training Center at Bainbridge, Maryland, served as *Ss*.[3] The *S* was seated approximately 10 feet from a black backdrop. During the experimental session the *S* was able to view only a single figure standing before the backdrop. Illumination was provided by four 60-watt bulbs mounted in reflectors which were focused on the figure being observed.

Two figures were observed by each *S*: (*a*) *Authority figure.* The observed individual was dressed in a white enlisted man's uniform. On his left sleeve was the rating badge of a 1st Class Boatswain's Mate and two hash marks. (It should be pointed out that the *S*s used had 1st Class Petty Officers as their Company Commanders.) (*b*) *Nonauthority figure.* The observed individual was dressed in a white enlisted man's uniform with the identifying marks of a recruit: canvas leggings or "boots" and the stripes of a Seaman Apprentice. (At Bainbridge all Seaman recruits wear the stripes of a Seaman Apprentice.)

Two different individuals played the roles of the authority and the nonauthority figure. Every effort was made to present these figures as genuine. Within each subsample of 12, one individual played the role or the authority figure 6 times and the nonauthority figure 6 times. The names assigned to the authority figure and to the nonauthority figure remained constant. The two individuals employed were approximately the same size, weight, and body build.

Each figure was viewed by the *S* through aniseikonic lenses in a series of 14 lenses ranging in power from 0.25 per cent to 4.50 per cent. Four series of lens presentations were given for each figure; 2 of the series were ascending, 2 were descending. The *S* attended a single experimental session which lasted approximately 50 minutes. The observed individual was visible to the *S* only during the actual presentation of each single lens power. After a lens had been inserted into the viewing apparatus, the *S* pressed a button which illuminated the figure standing in front of the backdrop. He then observed that figure for a period of 25 seconds, at which time an electronic timer shut off the light. Following each lens presentation the *S* reported the appearance of the observed individual for that particular lens presentation.

The *E* gave the following specific instructions: "Each time the light goes on you are to observe the appearance of the individual in the room as closely as possible. After the light goes off, I want you to tell me what he looked like when the light was on. If he looks the same as usual, I want you to tell me that. If he appeared to be changed in any way whatsoever, I want you to tell me that. I also want you to tell me in what way he changed. Now he may change in a number of ways. He may change in size; he may appear to tilt forward or backwards; he may change in his distance from you; he may change in specific parts of his body. Any one

[1] Now with the Armstrong Cork Company, Lancaster, Pa.

[2] The opinions or assertions contained in this article are the private ones of the authors and are not to be construed as official or reflecting the views of the Navy Department or the naval service at large.

[3] The authors wish to extend their sincerest thanks to Captain S. P. Sanford, MC, USN, and to Lt. W. B. Lyon, MSC, USNR, whose cooperation made this study possible.

or a combination of these changes may occur. In any event, please tell me after each lens presentation what his appearance was while the light was on."

The *E* recorded the point on the optical scale (lenses numbered 1 through 14) at which distortion of the observed individual was first reported when the series was ascending, or the point at which distortion was no longer reported if the series was descending. The starting point for any particular ascending or descending series was varied within conditions, but equated between conditions. Also, the order of presentation of the authority and the nonauthority figure, as well as the order of presentation of the ascending and descending series, was balanced within each subsample of 12 *S*s so as to cancel out any possible effects due to order of presentation.

RESULTS

For the total sample of 24 *S*s, the mean threshold for distortion for the Boatswain's Mate was 3.28 (*SD* 2.29); for the Seaman Apprentice 2.16 (*SD* 1.32). For 22 of the 24 *S*s, the threshold for distortion was higher for Boatswain's Mate than Seaman Apprentice. By sign test, this gives a two-tailed p value of less than .0001.

SUMMARY

As predicted from previous work on fear-evoking objects and resistance to perceptual distortion, Navy recruits viewing persons through aniseikonic lenses showed higher distortion thresholds when viewing an authority figure than when viewing a nonauthority figure.

REFERENCES

1. GILDER, R., THOMPSON, S. V., SLACK, C. W., & RADCLIFFE, K. B., JR. Amputation, body image, and perceptual distortion. *USN Med. Res. Inst. Rep.*, 1955. (Project NM 004 008.04.03.)
2. WITTREICH, W. J., & RADCLIFFE, K. B., JR. The influence of simulated mutilation upon the perception of the human figure. *USN Med. Res. Inst. Rep.*, June 1954. (Project NM 004 008.04.02.)

Received November 16, 1955.

A NOTE ON THE SOCIAL PERCEPTIONS OF AUTHORITARIANS AND NONAUTHORITARIANS

WILLIAM RABINOWITZ

Bank Street College of Education

SCODEL and Mussen (5) recently demonstrated that when authoritarian and nonauthoritarian persons, as measured by the F scale (1), were paired in a social situation, after which they estimated each others' F-scale responses, the nonauthoritarians were more accurate than the authoritarians. They interpreted their results as evidence that nonauthoritarians possess greater interpersonal sensitivity and insight than authoritarians. However, a more recent study by Scodel and Freedman (4) does not support this interpretation.

The Scodel-Freedman study essentially repeated the Scodel-Mussen design with one important change: authoritarians were paired with other authoritarians and nonauthoritarians were paired with other nonauthoritarians. In general the authoritarians tended to make accurate judgments about the F-scale scores of the other authoritarians with whom they had been paired, whereas the nonauthoritarians tended to make inaccurate judgments about the other nonauthoritarians. This result directly contradicts the assertion that nonauthoritarians are more insightful in interpersonal perception than authoritarians. Most revealing was the comparison that Scodel and Freedman provide of the two sets of results. In both studies, the F-scale estimates were highly related to the judge's own F-scale score but unrelated to the F-scale score of the partner with whom the judge was paired.[1] Authoritarians tended to estimate high F-scale scores for their partners, no matter what their actual score. Nonauthoritarians, however, tended to make more varied estimates of their partners, sometimes estimating high scores, sometimes low scores, but most often scores in the middle range.

Apparently, F-scale scores judged on the basis of a brief interaction reflect the assumptions of the judge, not the objective characteristics of the *S* judged. The findings of the Scodel-Mussen and Scodel-Freedman studies may arise from a difference in the beliefs held by authoritarians and nonauthoritarians about the F-scale responses of an "average" college student.

Gage and Cronbach, discussing conceptual and methodological problems in interpersonal perception, point out:

If we are asking a Judge to predict the response of an individual to a personality inventory, we . . .

[1] Since preparing this article, the writer's attention has been called to a study by Crockett and Meidinger (2) in which all combinations of subjects were used: authoritarian-authoritarian, nonauthoritarian-authoritarian, and nonauthoritarian-nonauthoritarian. Their results add further support to the assertion that the judge's estimate is not related to the actual F-scale score of his partner.

probably . . . want to measure "ability to predict how this individual deviates from the typical behavior of the particular group he belongs to." If accuracy is scored directly by comparing the prediction to the response of the individual, we are not distinguishing between two components which contribute to the Judge's success: his knowledge of the response that any individual in the subgroup is likely to give, and his knowledge of the way in which *this* individual deviates from the norm.

It is apparently desirable, when studying ability to predict at any one level, to obtain at least two scores: (*a*) ability to predict the typical behavior in the next-larger class to which Other belongs, and (*b*) ability to predict how Other deviates from the norm for this class (3, p. 417).

In neither of the studies under examination were *S*s asked to predict the F-scale responses typical in their college student group. This note describes a study undertaken without knowledge of Scodel and Freedman's results, in which *S*s with known F-scale scores marked the F scale as they believed a typical member of their college group would respond to it.

A group of 104 undergraduate college students (16 males, 88 females) were given a 28-item version of the F scale[2] twice: in the first administration, customary instructions were used, but in the second administration, *S*s were asked to give the responses which they believed a typical student at their college would give. The second administration immediately followed the first. No *S* had access to his original F-scale responses during the second administration.

Table 1 summarizes the results for the total group of *S*s, as well as the highest 20 per cent (authoritarians) and the lowest 20 per cent (nonauthoritarians). The authoritarians had F-scale scores between 104 and 136, and the nonauthoritarians had F-scale scores between 36 and 61. Analysis of these data suggests the following conclusions:

1. For the total group, the difference between the means of the first and second administrations is highly significant ($t = 3.54$, $p < .001$). There is apparently a general tendency for these *S*s to estimate higher F-scale scores for the typical college student than they themselves possess.

2. The mean estimate of the nonauthoritarians is approximately 38 points higher than their own mean, whereas the mean estimate of the authoritarians is approximately 9 points higher than their own mean. This difference is statistically highly significant ($t = 3.73$, $p < .001$). Apparently, the general tendency of the *S*s to estimate F-scale scores higher than their own is more marked among the nonauthoritarians than the authoritarians.

TABLE 1

F-SCALE SCORES OBTAINED UNDER TWO CONDITIONS OF INSTRUCTION

Instructions	Total Group ($N = 104$)		Nonauthoritarians ($N = 21$)		Authoritarians ($N = 21$)	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
Own attitude	83.1	23.4	51.3	8.0	116.0	9.3
Typical student	104.8	27.2	89.1	27.8	125.2	17.5

3. The mean estimate of the nonauthoritarians is only approximately 6 points higher than the actual total group average, whereas the mean estimate of the authoritarians is approximately 42 points higher than the actual group average. This difference is statistically highly significant ($t = 5.03$, $p < .001$). The nonauthoritarians are apparently more accurate than the authoritarians in estimating the true group average. Although this may reflect greater sensitivity to group opinion on the part of the nonauthoritarians, it seems more probable that it is a product of the general tendency of all *S*s to estimate scores higher than their own, which happens, in this case, to make the estimates of the nonauthoritarians more accurate.

4. The estimates of the nonauthoritarians are more variable than those of the authoritarians ($F = 2.51$; $.01 < p < .05$). Apparently the authoritarians share a more uniform conception of the typical student's responses than do the nonauthoritarians.

5. The correlation between the first and second administrations is highly significant ($r = .51$, $p < .001$), indicating the relative stability with which *S*s maintain their rank position in the two distributions. Apparently, responding as a typical college student, low-scoring *S*s tend to estimate low scores, and high-scoring *S*s tend to estimate high scores.

Although a direct comparison is precluded, the results summarized in Table 1, based on *S*s' estimates of the F-scale responses of a typical college student, are consistent with those of the Scodel-Mussen and Scodel-Freedman experiments. The experimental evidence demonstrates that judges who estimate the F-scale responses of others, after brief interpersonal contact, are unable to differentiate accurately among the others. The estimates depend more upon the F-scale score of the judge than that of the other. As in the cited experiments, the evidence of this study indicates than an *S*'s estimate of the F-scale responses of a typical student depends largely upon the *S*'s own

[2] The following F-scale items were used: Form 78, Items 6, 66, 67; Form 60, Items 22, 41, 50; Form 45 and 40, Items 1, 2, 6, 9, 12, 13, 18, 19, 21, 23, 26, 27, 29, 31, 33, 34, 37, 38, 41, 42, 43, 44 (1, pp. 226–227, 248–250, 255–257).

F-scale score. The combined data suggest two plausible hypotheses: (*a*) a judge who estimates the F-scale responses of an other, after brief interpersonal contact, does so by assuming the other is a typical member of the group to which he belongs, and (*b*) a judge's estimates of the F-scale responses of a typical member of a group depend more upon the judge's own F-scale responses than those characteristic of the group.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.
2. CROCKETT, W. H., & MEIDINGER, T. Authoritarianism and interpersonal perception. *J. abnorm. soc. Psychol.*, **53**, 378–380.
3. GAGE, N. L., & CRONBACH, L. J. Conceptual and methodological problems in interpersonal perception. *Psychol. Rev.*, 1955, **62**, 411–422.
4. SCODEL, A., & FREEDMAN, MARIA L. Additional observations on the social perceptions of authoritarians and nonauthoritarians. *J. abnorm. soc. Psychol.*, 1956, **52**, 92–95.
5. SCODEL, A., & MUSSEN, P. Social perceptions of authoritarians and nonauthoritarians. *J. abnorm. soc. Psychol.*, 1953, **48**, 181–184.

Received February 27, 1956.

A COMPARISON OF THE DRQ WITH RATINGS OF EMOTION[1]

FRANK AULD, JR. AND GEORGE F. MAHL

Yale University

THE Discomfort-Relief Quotient, or DRQ, was invented by Dollard and Mowrer (3, 4) as a method of measuring tension in written documents. To obtain the DRQ, a scorer classifies each unit of the written material as expressive of discomfort, relief, or neutral feeling, and then computes the ratio of discomfort units to the total of discomfort and relief units. Transcribed interviews, as well as communications originally made in written form, can be scored by this method.

The question naturally arises whether the quotient so obtained accurately measures the tension felt by the communicator. Meadow and his coworkers (8), believing that the DRQ might not accurately reflect psychological tension, compared the DRQ scores of 35 schizophrenic patients with a psychiatrist's ratings of the tension felt by these patients. Meadow *et al.* found a negligible and insignificant correlation between the DRQ scores and the ratings of tension. However, this result does not prove that the DRQ is invalid, because: (*a*) The ratings of tension were made in a different situation from that in which material for the DRQ scores was gathered; the lack of correlation may only show that patients behave differently in different situations. (*b*) Meadow *et al.* computed the DRQ erroneously, including all units in the denominator. (*c*) Since the subjects were schizophrenic, the results may only show that the DRQ is invalid for psychotics, who are known (2) to verbalize their feelings less accurately than normals and neurotics.

The aim of the present study is to appraise the validity of the DRQ as a measure of tension, using as subjects a group consisting mostly of neurotics, and obtaining the ratings of tension from the same situation that provides the material for the DRQ scores.

METHOD

Subjects. The *S*s are 39 persons who either applied for treatment in the Psychiatric Outpatient Clinic of the New Haven Dispensary or entered the Yale Psychiatric Institute, a private psychiatric hospital, as patients. Of the 39 patients, 31 are outpatients and 8 are inpatients. Twenty-two of the *S*s were diagnosed as psychoneurotic, 10 as suffering from character disorders, and 7 as psychotic.

Procedure. Three interviewers—a psychologist and two psychiatric residents—took turns interviewing the *S*s, one of the team doing the interviewing while the others monitored and observed through a one-way mirror.[2] For the outpatients these were initial interviews, conducted in order to permit a diagnostic formulation; for the inpatients these are special research interviews. Immediately after each interview all members of the team made independent ratings of the patient's behavior, on a five-point scale, on each of three variables: anxiety, hostility, and dependence. All interviews were recorded on magnetic tape, were later carefully transcribed, and were divided into sentence units according to rules developed by Auld, Dollard, and White (1). The two authors and a secretary scored the interviews according to the DRQ method, one author (G.F.M.) scoring the entire series,

[1] The authors wish to thank the National Institute of Mental Health, U. S. Public Health Service, for financial support of this investigation, through grant M-648, "Development of Quantitative Methods for the Study of Psychotherapy" (Professor John Dollard, principal investigator), and grant M-1052, "The Patient's Language as Expressive Behavior" (George F. Mahl, principal investigator). The authors are also grateful for financial aid from the Foundations Fund for Research in Psychiatry.

[2] These interviews and ratings were obtained for other purposes by Drs. George Andrews, George F. Mahl, and Louis Micheels. The facilities for the recordings and observation of these interviews have been described elsewhere (7). Additional information about the interviews is presented in a previous paper (5).

TABLE 1
RELIABILITY OF RATINGS OF EMOTION

Variable	Interjudge Reliability Coefficients		
	A with B	A with C	B with C
Anxiety	.68****	.22	.42***
Hostility	.61****	.46***	.47***
Dependence	.47***	.37**	.28*

* Significant at .10 level.
** Significant at .02 level.
*** Significant at .01 level.
**** Significant at .001 level.

TABLE 2
CORRELATIONS OF COMPOSITE RATINGS WITH DRQ

Variable	Correlation Coefficients		
	Men ($N = 13$)	Women ($N = 26$)	Total Sample
Anxiety	−.11	.33*	.27*
Hostility	.02	.15	.12
Dependence	.24	.37*	.31*
Anxiety + Hostility + Dependence	.15	.41**	.35**

* Significant at .10 level.
** Significant at .05 level.

TABLE 3
CORRELATIONS OF COMPOSITE RATINGS WITH DRQs OF TWO SCORERS

Variable	Correlation Coefficient*	
	DRQ of F.A.	DRQ of G.F.M.
Anxiety	.08	.17
Hostility	.10	.04
Dependence	.24	.36
Anxiety + Hostility + Dependence	.21	.28

* With a sample of only 21 cases, none of the reported coefficients is statistically significant at the .10 level.

the other (F.A.) scoring 21 interviews, and the secretary scoring 10 interviews.

It was planned to have the secretary score all the interviews and to use her DRQ scores in the analysis of results, provided that she should score reliably. When it turned out that her scores had a correlation of only .51 with G.F.M.'s and .53 with F.A.'s, the authors decided to use G.F.M.'s scores instead.

RESULTS

Reliability. The agreement among judges for the ratings of anxiety, hostility, and dependence is reported in Table 1. To obtain the most reliable single measure of anxiety and hostility, we added A's and B's ratings, ignoring C's. To obtain the most reliable measure of dependence, we took the sum of all three ratings. As estimated by the Spearman-Brown formula the reliability of the anxiety composite is .81, of the hostility composite, .76, and of the dependence composite, .62.

Reliability of the division of interviews into sentences was established by comparing the secretary's unitizing with that of the authors. There was agreement on placement of more than 95 per cent of the unit markings.

The correlation between the authors' DRQ scores of the 21 hours scored by both was .90, which is statistically significant at the .001 level.

Correlation of ratings with DRQ. Table 2 shows the correlations of the composite ratings with the DRQ, for men and for women separately and for the total sample. Taken all in all, there is a slight positive relationship between the ratings on these variables and the DRQ.

The correlations between the ratings and the DRQ are consistently higher for the women than for the men. The probability is about .22, as figured by Stouffer's method of combining probabilities (9), that three such differences between correlations for men and for women as we have observed would occur simultaneously by chance. A study by Mahl (6) of the correlation between disturbances in speech and the anxiety ratings has revealed a similar and highly significant sex difference. The meaning of these sex differences is obscure at the present time.

Possibility of contamination. Since one of the raters was also the person whose DRQ scores were reported above, it is possible that any correlation between the ratings and the DRQ scores is spurious. To exclude this possibility we separately analyzed the 21 interviews which the other author had scored. The results of this analysis are presented in Table 3. No conclusion of this study would be changed if F.A.'s scores rather than G.F.M.'s were used.

DISCUSSION

It does not seem likely that the small size of the correlations between the global ratings and the DRQ can be accounted for by poor reliability of the measures. When the correlations are corrected for attenuation they are still of unimpressive size. To the degree that the global ratings are themselves valid, therefore, the failure of the DRQ scores to correlate with them more highly should discourage the use of the DRQ as a measure of individual differences in anxiety, hostility, or dependence.

It must be noted that this conclusion applies, strictly, only to use of the DRQ as a measure of *individual differences* in tension. The findings do not tell whether a change in the DRQ within an interview series of a single patient would validly measure changes in the tension felt by the patient.

The reader's attention is directed to the high reliability of the DRQ scores. Whereas interjudge

reliabilities ranged from .22 to .68 for the global ratings of emotion, the interjudge reliability of the two authors' DRQ scores was .90. The superior reliability of the DRQ procedure suggests that one might profitably seek to increase the reliability of measures of anxiety, hostility, dependence, and the like by adopting scoring procedures patterned on the DRQ. Perhaps such procedures, based on content-analysis methods and including behavior items and vocal as well as verbal cues, could measure essentially the same variables as the global judgments but do so with greater precision.

SUMMARY

In order to appraise the validity of the DRQ as a measure of tension, the authors compared DRQ scores of 39 psychiatric interviews with global ratings of the same interviews on anxiety, hostility, and dependence. There were small positive correlations between the global ratings and the DRQ, these correlations tending to be higher for the women than for the men. If the global ratings can be accepted as adequate measures of tension, the DRQ does not measure individual differences in tension very well. However, one must keep in mind that this study focuses on measurement of individual differences in tension and is not immediately pertinent to use of the DRQ as a measure of changes in tension within a series of interviews with a single patient.

REFERENCES

1. AULD, F., DOLLARD, J., & WHITE, ALICE M. A method of dividing interviews into independent-clause units. Paper read at East. Psychol. Assoc., Philadelphia, April 16, 1955.
2. DOLLARD, J. The psychotic person seen culturally. *Amer. J. Sociol.*, 1934, **39**, 637–648.
3. DOLLARD, J., & MOWRER, O. H. Progress report on casework research. New York: Community Service Society, 1943.
4. DOLLARD, J., & MOWRER, O. H. A method of measuring tension in written documents. *J. abnorm. soc. Psychol.*, 1947, **42**, 3–32.
5. LAMB, R., & MAHL, G. F. Manifest reactions of patients and interviewers to the use of sound recording in the psychiatric interview. *Amer. J. Psychiat.*, 1956, **112**, 731–737.
6. MAHL, G. F. Disturbances in the patient's speech as a function of anxiety. Paper read at East. Psychol. Assoc., Atlantic City, March 23, 1956.
7. MAHL, G. F., DOLLARD, J., & REDLICH, F. C. Facilities for sound recording and the observation of interviews. *Science*, 1954, **120**, 235–239.
8. MEADOW, A., GREENBLATT, M., LEVINE, J., & SOLOMON, H. C. The Discomfort-Relief Quotient as a measure of tension and adjustment. *J. abnorm. soc. Psychol.*, 1952, **47**, 658–661.
9. MOSTELLER, F., & BUSH, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 289–334.

Received June 29, 1956. Prior publication.

A SIMPLE DEVICE FOR OBTAINING CERTAIN VERBAL ACTIVITY MEASURES DURING INTERVIEWS[1]

STANISLAV V. KASL AND GEORGE F. MAHL

Yale University

DURING the past decade and a half there have been many objective studies of interview transactions. The vast majority of these have consisted of verbal content analyses, i.e., the reliable description in quantitative terms of "what" the participants have said. The "how" (6), or the "vocal" (7), aspects of the communication have received limited attention in spite of the common knowledge that they are just as important as the verbal content, and often more so.

One particular class of the "how" of verbal activity includes those features of speech that are exclusively behavioral (i. e., without conventional semantic function), such as the amount of talking by the participants in standard time intervals, the rate of talking when actual speaking occurs, the amount of silence, interruptions, etc. Any extensive study of such verbal activity, however, requires some instrumentation beyond a stop watch. Chapple (1, 2) devised his Interaction Chronograph and Verzeano and Finesinger (9) their Automatic Speech Analyzer for such purposes. Matarazzo *et al.* (5) give a comprehensive picture of the functioning of the Chapple Interaction Chronograph, the data obtained with it, and of studies based on its use. Applications of the Verzeano-Finesinger device are found in papers by Verzeano (8), and Lorenz and Cobb (3).

During the course of investigating the expressive aspects of the patient's language in psychotherapy, it became necessary to determine the duration of time spent talking by the participants, and the amount of time the patient was silent during successive two-minute intervals of individual interviews as well as during larger and variable time intervals. An instrument for these relatively restricted purposes does not have to be as elaborate as the Interaction Chronograph or Speech Analyzer which yield additional data, and yet it requires

[1] This study is a fragment of a research program supported by USPHS Grant M-1052, The Patient's Language as Expressive Behavior (George F. Mahl, principal investigator), and by the Foundations' Fund for Research in Psychiatry.

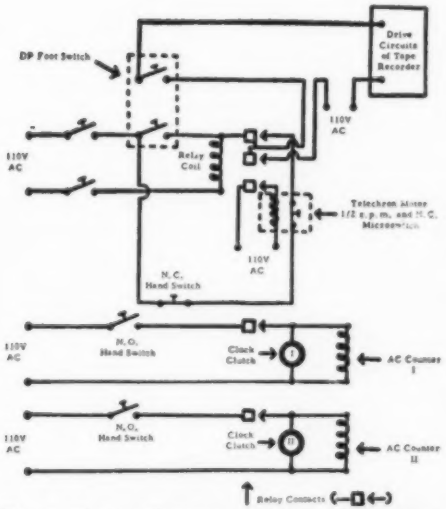

FIG. 1. CIRCUIT DIAGRAM OF TIMING INSTRUMENT

certain features not found in them. Consequently, one of the writers (G.F.M.) devised the simple instrument that is described in this paper. This is essentially a development of Chapple's original idea for a special purpose. The instrument is inexpensive, has been of practical value, and its use yields reliable data. Others may also find it to be helpful.

Figure 1 contains the circuit diagram. The instrument is for use with tape recordings of interviews and playback equipment that is started and stopped with electronic rather than mechanical controls. How the instrument is connected with the electronic controls of the driving mechanism of the playback equipment may vary in detail with the particular machine that is used. Since there is nothing special about the parts used in the apparatus, no detailed specifications are presented.

An observer operates the apparatus while listening to the recording through earphones. In obtaining the successive two-minute measures, the instrument functions as follows. The momentary closure of the observer's foot-switch starts the tape recorder, a telechron motor geared for ½ r.p.m., and closes the relay switches in series with the clutches of two accumulating clocks (Springfield timers) and the coils of two counters. The operator may then release his foot and the various elements will still be activated since the relay is wired as a holding relay. Now the observer operates the two hand-switches (microswitches equipped with a metal button on the switch arm for sensitive finger operation) that are also in series with the clock clutches and the counter coils. When he now presses them each time the patient or the therapist talks, the appropriately assigned clock will run for as long as the key is depressed and the counter will make one count per closure and opening of the switch (this is called an "act"). The "Normally Closed Microswitch" is opened when the telechron motor completes one revolution in two minutes. This instantly releases the relay, opening all the relay contacts so as to stop the tape recorder and inactivate the clocks and counters. The clocks and counters are then read and zeroed. Starting again with a momentary closure of the foot-switch, the same sequence is now repeated for the next two minute. By using this apparatus the raw data for a 50–60 minute interview, consisting of the number of seconds of talking and the number of "acts" for each participant in each two-minute interval, can be obtained in about one and a half hours by an experienced operator. Practice runs on 10 to 15 interviews give an inexperienced person sufficient training.

If one wishes to use a fixed time segment other than of two minutes duration, all that is required is to use a telechron motor that is geared to make 1 r.p.m. in the preferred time interval.

For some purposes it is desirable to obtain the activity measures for variable time segments, e.g. during a period when the patient is talking about various specific topics, or when he is engaged in a particular defense, etc. This is done with the present device by short circuiting the microswitch

TABLE 1

TEST-RETEST AND INTEROBSERVER CORRELATIONS OF TWO-MINUTE ACTIVITY MEASURES IN 5 UNSELECTED INTERVIEWS*

Activity Measure	Observers Compared	Product Moment Correlations $\bar{X}$	Range
Seconds of Patient Talk.	A-A	.98	.957–.992
	B-B	.97	.923–.986
	A-B	.97	.940–.986
Seconds of Therapist Talk.	A-A	.99	.972–.997
	B-B	.98	.965–.986
	A-B	.99	.972–.996
N Patient "Acts"	A-A	.88	.455–.954
	B-B	.78	.602–.879
	A-B	.76	.566–.910
N Therapist "Acts"	A-A	.94	.928–.954
	B-B	.92	.796–.964
	A-B	.94	.882–.971
Silence Quotient	A-A	.97	.943–.990
	B-B	.91	.817–.949
	A-B	.94	.896–.973

* These correlations reflect the observations of Jon Eklund and Stanislav Kasl.

Observer A's retests made 2–4 weeks after initial test.

Observer B's retests made 6 months after initial runs of these 5 and 34 other interviews with no intervening practice on other material.

operated by the telechron motor. The tape is set to start with the onset of the particular phase. The operator then starts with the momentary closure of the foot-switch. But now, with the telechron microswitch shorted, the apparatus will run until the "normally closed hand-switch" in Fig. 1 is opened. We have used a second observer to monitor the tape along with the operator and to open this hand-switch at the end of the particular phase. The operator is thus left free to attend only to the ongoing verbal behavior in the tape, so that he is not distracted by having the additional task of "listening for" the end of the phase.

Three people have learned independently to operate the apparatus, and in each case they developed similar habits to maximize their efficiency and accuracy. It was felt necessary, for example, to concentrate on the speech, yet pay no attention to its actual content or meaning. Eyes were fixed on some neutral object in the room or on the hand of the Springfield timer, perhaps to restrict the range of attention. The resulting set seemed to facilitate the observers' responding to the speech as an irregular, intermittent auditory stimulus, rather than to such linguistic units as phrases and sentences.

We have used the activity measures in two ways. The number of seconds of talking has been used with word counts to obtain "rates of talking when talking." Also a Patient Silence Quotient has been determined for successive two-minute intervals of therapeutic interviews, and in various motivational phases of the interview. The Patient Silence Quotient[2] is defined as:

$$\text{Silence Quotient} = \frac{N \text{ seconds of silence}}{N \text{ seconds available to patient to talk}}$$

$a = N$ seconds in the segment (120″ e.g. in 2′)
$b = N$ seconds talk by therapist
$c = N$ seconds talk by patient
N seconds of silence $= a - (b + c)$
N seconds available to patient $= a - b$.

The count of "acts" has not yet been used. But it is possible that they will help provide a measure of the speaker's fluency, independent of the Silence Quotient. Table 1 contains test-retest and inter-observer correlations for the various activity measures in two-minute intervals of 5 unselected interviews. The *N*s ranged from 25 to 30 two-minute intervals.

[2] It can be seen from the definition of the Silence Quotient that sometimes brief hesitations, "naturally" attributable to the therapist, that is, in which the therapist gives every indication of continuing his talk but does not do so for a brief moment, are scored as "time available to patient." The amount of time that is thus erroneously treated is regarded as negligible.

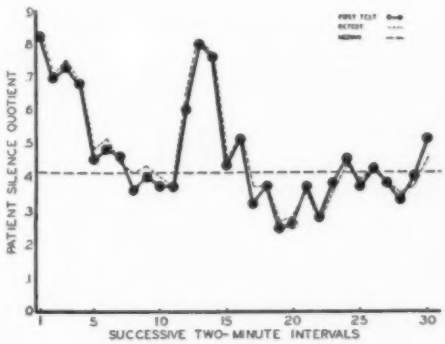

FIG. 2. TEST-RETEST PATIENT SILENCE QUOTIENTS IN SUCCESSIVE TWO-MINUTE INTERVALS OF A THERAPEUTIC INTERVIEW. MEDIAN IS OF FIRST SET OF MEASURES

Figure 2 is a graph of the Silence Quotient in successive two-minute intervals of a therapeutic interview. This graph illustrates the test-retest reliability, as well as the clear-cut depiction of behavioral change during the course of an interview that is possible by means of the measurement of "how" the patient talks. A previous paper (4) has dealt with some of the broader implications of such data.

REFERENCES

1. CHAPPLE, E. D. Measuring human relations: an introduction to the study of interaction of individuals. *Genet. Psychol. Monogr.*, 1940, **22**, 3–147.
2. CHAPPLE, E. D. The interaction chronograph: its evolution and present application. *Personnel*, 1949, **25**, 295–307.
3. LORENZ, M., & COBB, S. Language behavior in manic patients. *Arch. Neurol. Psychiat.*, 1952, **69**, 763–770.
4. MAHL, G. F. Disturbances and silences in the patient's speech in psychotherapy. *J. abnorm. soc. Psychol.*, 1956, **53**, 1–15.
5. MATARAZZO, J. D., SASLOW, G., & MATARAZZO, R. G. The interaction chronograph as an instrument for objective measurement of interaction patterns during interviews. *J. Psychol.*, 1956, **41**, 347–367.
6. REICH, W. On character analysis. In R. Fliess (Ed.), *The psychoanalytic reader*. New York: International Universities Press, 1948.
7. SULLIVAN, H. S. *The psychiatric interview*. New York: Norton, 1954.
8. VERZEANO, M. Time-patterns of speech in normal subjects. *J. Speech Hearing Disorders*, 1950, **15**, 197–201.
9. VERZEANO, M., & FINESINGER, J. E. An automatic analyzer for the study of speech in interaction and in free association. *Science*, 1949, **110**, 45–46.

Received July 24, 1956. Prior Publication.

You are urged to read

THE AMERICAN JOURNAL OF SOCIOLOGY

Begin with September, 1956, an issue devoted to

"THE INTERVIEW IN SOCIAL RESEARCH"

TABLE OF CONTENTS

Of Sociology and the Interview: An Editorial Preface. By MARK BENNEY and EVERETT C. HUGHES, *University of Chicago*. The interview, the art of sociological sociability, is both a scientific tool and a part of the very subject matter of sociology. In this new form of social rhetoric the part of the interviewer is already conventionalized, and now laymen are learning how to play their roles as respondents.

Age and Sex in the Interview. By MARK BENNEY, DAVID RIESMAN, and SHIRLEY A. STAR, *University of Chicago*. The varying combinations of age and sex of interviewer and interviewed have profound effects on the responses.

Role Relationships and Conceptions of Neutrality in Interviewing. By LEWIS ANTHONY DEXTER, *Chevy Chase, Maryland, and Belmont, Massachusetts*. The interviewer must take into account in his phraseology and gestures the fact that the respondent identifies him as belonging to a certain class and cherishing its values.

Dimensions of the Depth Interview. By RAYMOND L. GORDEN, *Antioch College*. Interviews with individuals involved in disaster shed light on problems of eliciting emotionally charged information and on differences in sociological and therapeutic procedures.

The Dynamics of Information Interviewing. By THEODORE CAPLOW, *University of Minnesota*. The interview is an unusually gratifying conversation in which the interviewer must be neutral and yet encouraging.

Role-playing in Survey Research. By HOWARD STANTON, KURT W. BACK, and EUGENE LITWAK, *University of Puerto Rico, University of North Carolina, and University of Chicago*. By staging a sociodrama with respondents, the research worker may learn how they behave in situations where they are acutely responsive to the reactions of others.

The Respondent Reports on the Interview. By CHARLES F. CANNELL and MORRIS AXELROD, *University of Michigan*. Most laymen find being interviewed so interesting that they will answer even embarrassing personal questions fully and frankly.

A Classification of Biased Questions. By EUGENE LITWAK, *University of Chicago*. From a systematic analysis of ambiguity and distortion, criteria are derived for more precise wording of questions.

Interviewing Frenchmen. By DANIEL J. LERNER, *Massachusetts Institute of Technology*. Research workers must take account of the conventions which maintain social distance in French conversation and determine what role is feasible to the interviewer.

FIELD NOTES

Interviewing Negro Pentecostals. By NICHOLAS VON HOFFMAN and SALLY W. CASSIDY, *University of Chicago*.

Interviewing Water Dowsers. By EVON Z. VOGT, *Harvard University*.

Interviewing Medical Students. By HOWARD S. BECKER, *University of Chicago and Community Studies, Inc., Kansas City, Missouri*.

Interviewing Homosexuals. By MAURICE LEZNOFF, *University of Chicago*.

Multilingual Interviewing in Israel. By HAIM BLANC, *Hebrew University, Jerusalem*.

The Unidentified Interviewer. By HENRY W. RIECKEN, *University of Minnesota*.

The Personal Identification of the Interviewer. By PETER KONG-MING NEW, *Community Studies, Inc., Kansas City, Missouri*.

Repeated Interviewing. By MURRAY WAX and LEO J. SHAPIRO, *Chicago*.

NEW subscribers are offered a special rate of $4.50 for 8 issues; September, 1956 to November, 1957.
Order from the University of Chicago Press,
5750 Ellis Avenue,
Chicago 37, Illinois

PSYCHIATRY

JOURNAL FOR THE STUDY OF INTERPERSONAL PROCESSES

Contents of No. 3, Vol. 19, August 1956

DONALD HORTON AND R. RICHARD WOHL: *Mass Communication and Para-Social Interaction.*

DAVIDE LIBERTANI: *Symbiotic Identification in Schizophrenia.*

JOHN H. WEAKLAND: *Orality in Chinese Conceptions of Male Genital Sexuality.*

ELAINE CUMMING, I. L. W. CLANCEY, AND JOHN CUMMING: *Improving Patient Care through Organizational Changes in the Mental Hospital.*

EDWARD STAINBROOK AND MURRAY WEXLER: *The Place of the Behavioral Sciences in the Medical School.*

JOHN L. CAMERON, THOMAS FREEMAN, AND ANDREW MCGHIE: *Clinical Observations on Chronic Schizophrenia.*

DONALD L. BURNHAM: *Misperception of Other Persons in Schizophrenia.*

EDITORIAL NOTES.

BRIEF COMMUNICATIONS.

BOOK REVIEWS.